Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
durch die Wirtschaftswissenschaftliche Fakultät
der Westfälischen Wilhelms-Universität Münster

Bank Efficiency Estimation

Methodology and the Problem of Adequation

Sebastian D. Tente

Wirtschaftswissenschaftliche Fakultät

der Westfälischen Wilhelms-Universität

Universitätsstraße 14-16, 48143 Münster

Dekan: Prof. Dr. Thomas Apolte

Erstgutachter und Betreuer: Prof. Dr. Andreas Behr

Zweitgutachter: Prof. Dr. Andreas Pfingsten

Tag der mündlichen Prüfung (Disputation): 23. November 2010

Es gibt keine schlechthin 'objektive' wissenschaftliche Analyse des Kulturlebens oder – was vielleicht etwas Engeres, für unseren Zweck aber sicher nichts wesentlich anderes bedeutet – der 'sozialen Erscheinungen', unabhängig von speziellen und 'einseitigen' Gesichtspunkten, nach denen sie – ausdrücklich oder stillschweigend, bewußt oder unbewußt – als Forschungsobjekte ausgewählt, analysiert und darstellend gegliedert werden.

Max Weber, Die 'Objektivität'

Contents

List of Figures

List of Tables

Chapter 1

1 Outline of the research project

1.1 Starting Point

This thesis addresses various topics on bank efficiency estimation. Roughly speaking our endeavours diverge into two lines of research: First, an argumentative framework discussing specific issues when setting up models for bank efficiency estimation. This is what we call the problem of adequate modelling in bank efficiency. And second, a methodological part covering innovations in econometric efficiency estimation.

Basically, we do not consider it appropriate to mix up both issues, i.e. to deal with methodological innovations and the adequation problem in a single study: Readers familiar with contemporary literature on bank efficiency[1] expect the application of certain traditional schemes when a survey announces to deal with the assessment of the performance of financial institutions. Even though we remain critical of some aspects of efficiency measurement especially in banks, we will stick to these well-established models, though, when focusing on the application of methodological innovations. First, this facilitates the comparability of our formal results with those of the traditional bank efficiency literature, and second, the methods we put forward are universally suitable in the course of any efficiency analysis. The problem of adequate modelling in bank efficiency has been split off. Consequently, this survey is composed of several modules, each of which can be worked through separately. As a matter of fact, occasional intersections of the basic elements of efficiency estimation are not always avoidable.

In the next two subsections 1.1.1 and 1.1.2 we will introduce our two lines of research in more detail. To conclude this introductory presentation of our research project, section 1.2 contains brief summaries

[1] For a recent survey, see Berger (2007), and also Berger et al. (1993).

of our projects that are documented in detail in the respective chapters.

1.1.1 Adequate modelling

In the course of the public discussion on the economic and financial crisis banks are given the characteristic of 'systemic relevance'. According to the German 'Bundesanstalt für Finanzmarktstabilität' (SoFFin), systemic relevance can be defined as follows[2]:

Systemic relevance arises when structures in the economy are tightly interwoven and cannot be easily broken up or replaced, as it is the case in the financial sector. In such structures the collapse of a company (e.g. a bank) causes not only other companies within the structure to collapse, but the entire economic system. To prevent a global breakdown of the economy, the government has to intervene. The costs for society are generally lower than the economic costs in the case of non-intervention.

So we see that in modern economies banks provide the basic infrastructure for capital transactions as well as the capital resources of the non-monetary sectors. According to a traditional view on bank activities, financial institutions might be reduced to the task of collecting deposits from surplus-spending units and lending those funds to deficit-spending units (typically corporate clients) (Sealey and Lindley, 1977). The transformation of many small short-term savings into a large long-term exposure yields a positive interest margin for the banks. Porter (1961) calls this the very essence of banking, which is to *borrow short and lend long*.

Even modern empirical bank efficiency estimation is based on this traditional view of financial intermediation. And the extent of the available literature adequately adresses the outstanding importance of an objective assessment of banks' productivity and efficiency. As the performance of single banks is rather uninteresting to the researchers, the focus of most studies lies on the entirety of banks, usually equated with the 'financial system': Especially the consolidation of the banking landscape in the European Union, the liberalization of markets in Central and Eastern Europe, as well as in the Asian emerging markets (e.g. India)

[2] This is a translation following the frequently asked questions on the SoFFin website http://www.soffin.de.

drive the need to 'measure' the impact of changing political and ecomonic conditions on banks' efficiency.

Nevertheless we see that banks are the main causes of the actual financial crisis. Obviously, banks' ability to manage portfolios of derivative products basically failed. This has not been foreseen in empirical bank efficiency literature, because the applied intermediation theory does not capture these aspects. As a result we learn about 'financial systems' assumed to be highly efficient but which, as a matter of fact, only survive with the aid of government interventions.

Following Grohmann (1988), we would like to call this discrepancy between empirical results on the one side and observations in the real world on the other side a 'problem of adequate modelling'. The discussion of statistical adequation in the context of the empirical bank efficiency estimation thus constitutes the economic subject matter of our research project. It is our aim to initiate a discussion reconsidering the contemporary modelling of production processes in banking.

1.1.2 Methodological issues

The second focus of our research project is on the implementation of methodical approaches to efficiency measurement by means of Stochastic Frontier Analyses (SFA). The estimation of SFA-models differs from classical linear models with respect to the assumptions imposed on the noise term. In particular, we expect banks to fall short of the optimal production level not only because of random environmental influences, but also due to managerial inefficiency. Commonly, we assume that the realization of firm-specific inefficiency is a random variable with an asymmetric distribution. This makes sense as inefficiency can only exert a negative influence on production and a positive influence on costs, respectively. The assumption of a composed error term certainly prohibits the estimation of the model using Ordinary Least Squares (OLS). It is therefore advisable to resort to Maximum Likelihood Estimation (MLE), or to correct a biased OLS-estimation by means of the Method of Moments (MOM).

After estimation of a Stochastic Frontier the question remains in how far the deviation between the actual production of a bank and the respective production frontier can be traced back to random noise or inefficiency. Dealing with cross-sectional data this question cannot be definitely answered and finally depends on the underlying assumption regarding the specific random distribution of inefficiency. As nobody ever observed inefficiency the choice of a one-sided distribution seems to be arbitrary (Førsund et al., 1980, p.11). Unfortunately, this choice conversely determines the estimated inefficiency. This is what one might call a tautology.

The solution to this problem may be to abstain from the application of SFA. Alternative methods are the Data Envelopment Analysis (DEA) as well as the Quantile Regression (QR). Both methods basically abandon the existence of environmental random influences on production. In consequence, the declared inefficiency is usually higher in DEA and QR than in SFA. Not least for this reason SFA is still the preferred method for efficiency estimation. Moreover, the above-mentioned problem of tautology can be overcome in the case of panel data, i.e. the existence of multiple observations per bank.

Since the pioneering work of Aigner et al. (1977), Meeusen and van de Broeck (1977) many outstanding econometricians developed advanced methods in SFA. Besides the rather uninteresting problem of implementing new inefficiency distributions, the main focus lies in the modelling of time-variant inefficiency when dealing with panel data. Ideally, the trend of firms' efficiency can be estimated as a parametric function simplifying the assessment of managerial learning effects or the success of economic policy measures.

In the framework of our survey we cannot pursue the aim of developing pioneering new methods in SFA. We rather observe that many of the approaches that have already been introduced in literature have never been applied in empirical studies. Indeed, this seems to be related to the fact that standard statistical software packages presently only allow estimation of basic SFA models.

Another approach to frontier estimation nearly unnoticed to date is the Quantile Regression (Koenker and Bassett, 1978, Koenker, 2005). We already mentioned that due to the rejection of a random noise term QR shares some disadvantages with DEA; but otherwise, QR provides some very appealing opportunities remedying other shortcomings.

We are about to show that it is possible to formally adapt complex models in SFA and QR to specific empirical problems with panel data. Basically, the optimization problems can be solved numerically with statistical software, e.g. the **R** Environment (R Development Core Team, 2010). Nonetheless, we try to never lose track of econometric standard methods, both robust and available to the wide range of authors with mainly empirical research problems.

1.2 Our contributions

1.2.1 Discussion of the empirical assessment of bank efficiency in the frame of a scheme of adequation

The empirical assessment of bank efficiency is a well-established branch of economic research today. Against the backdrop of the current economic crisis, however, the fact has become evident that no link can be established between the fundamental failure of the financial system and the results of these studies.

The discrepancy would seem to suggest that efficiency is assessed in a sector of bank activity which is not (no longer) relevant for the success of the institutions. In order to verify this supposition we have recourse to the scheme of 'statistical adequation' according to Grohmann (1988) to ascertain to what extent and at which place the activities of banks as laid down in the studies might deviate from the readers expectations.

As a result we observe that there exists a considerable adequation gap between the claim of the studies to make statements on the efficiency of financial systems on the one hand and the devices applied on the other hand.

1.2.2 Methods of cross-sectional stochastic frontier analyses[3]

The stochastic frontier analysis (Aigner et al., 1977, Meeusen and van de Broeck, 1977) is widely used to estimate firm-specific efficiency scores. The fundamental difference to ordinary least squares is the introduction of a two-part error term consisting of a noise and an inefficiency

[3] A first draft of this topic can be found in Behr and Tente (2008).

term. Most often the assumption of a half-normal distributed inefficiency term is applied, but other distributions are also discussed in the relevant literature. The natural estimation method seems to be Maximum Likelihood (ML) estimation because of the parametric assumptions. But simulation results obtained for the half normal model indicate that a method of moments approach (MOM) (Olson et al., 1980) is superior for small and medium sized samples in combination with inefficiency not strongly dominating noise (Coelli, 1995). In the frame of this paper we provide detailed simulation results comparing the two estimation approaches for both the half-normal and the exponential approach to inefficiency.

Based on the simulation results we obtain decision rules for the choice of the superior estimation approach taking into account the sample size and the efficiency to noise ratio. Both estimation methods, ML and MOM, are applied to a sample of German commercial banks based on the Bankscope database for estimation of cost efficiency scores.

1.2.3 Cost efficiency trends in European and US banking

The assessment of bank efficiency in the course of empirical studies dealing with international bank balance sheet and income statement data has been a frequently discussed field of application in recent decades. Based on firm-specifically estimated efficiency scores, the mean efficiency scores per country are ranked against each other to finally gain insight into the efficiency of national banking systems.

When ranking the banking systems of several countries against each other, it is essential to set up a common internatinal benchmark for all banks. With reference to the findings of Berger (2007), it is advisable furthermore to control for the national-specific (legal and social) environment the banks operate in via selected macroeconomic variables.

As most studies deal with panel data covering several years, the Stochastic Frontier Analysis SFA seems to be the preferred choice, as it allows to incorporate certain functional patterns of efficiency change over time. It is by now common pratice to control for country-specific heterogeneity in the macroeconomic environment mentioned above, but, to our best knowledge, there is not a single study allowing for country-specific heterogeneity in the efficiency trend parameters.

For this reason we put forward two modified methods based on mainstream fixed effects and random effects approaches to SFA. Concerning the fundamental understanding of efficiency of financial intermediation processes, we followed the most recent studies, e.g. Girardone et al. (2009). As a result we show that efficiency trends are in no way homogeneous across international banking systems.

1.2.4 Bank efficiency estimation based on quantile cost functions with fixed effects

Cross-sectional Quantile Regression (QR) in bank efficiency estimation has recently been proposed by Behr (2010). First, QR overcomes the limitations of Stochastic Frontier Analyses, because the estimated slope parameters of cost/production functions do not necessarily have to resemble the conditional mean parameters (OLS). Second, QR is robust against data outliers and violated assumptions regarding the noise term.

While first approaches were based on cross-sectional data, we venture to put forward an extended model to measure firm-specific inefficiency with panel data. Based on Koenker (2004), we estimate multiple conditional quantile cost functions with fixed effects which indicate time-specific location shift effects. This allows the cost level to vary from year to year, while the slope parameters remain constant over time. We are about to explore the new opportunities in the course of efficiency estimations. In addition to obvious distance-to-frontier measures of inefficiency, we propose a sector concept assigning efficiency classes to every firm.

Chapter 2

Preliminary reflections on efficiency and banks

2.1 Introduction

Before expatiating on the special topics of bank efficiency estimation in the following chapters we would like to acquaint the reader with the basic tools and the mindset necessary to understand the possibilities and limitations of bank efficiency estimation. As 'efficiency' is a determinant of performance, viz the degree of utilization of available resources[1], some preliminary reflections on various aspects of bank performance (section 2.2) – beyond the usual scope of the relevant empirical literature – are meant to prepare the reader for our critical appreciation of the applied adequation. To subsequently introduce our technical line of research, section 2.3 lays down the reasons why empirical literature rather relies on cost/profit frontier techniques than on traditional accounting ratios. Finally, section 2.4 provides an illustrative example of how we deal with the underlying Bankscope database. It is obvious that numerous problems we encounter in bank efficiency estimation can be traced back to the poor quality of publicly available data and the frequent occurence of data cleansing procedures.

2.2 Aspects of bank performance

As we already mentioned in the outline of our research project bank efficiency estimation rests upon rather entrenched structures targeted on costs of financial intermediation on the one hand (the 'scientific

[1] Fried et al. (2008) enumerates other determinants of performance, namely the production technology, the operating environment and the scale of operations.

approach') or shareholders' returns on the other hand (the 'practitioners' approach'). So literature database queries for the term 'bank efficiency' usually lead us to studies dealing with concepts of elaborately estimated cost functions or simple accounting ratios. At this point, we should not like to expatiate on the term of 'efficiency' yet. A definition of the term 'efficiency' will be developed in the subsequent section. For the moment, we will rather employ the term 'efficiency' as synonymous with the generic term 'performance'.

In the course of the studies dealing with bank efficiency, the reader gets no answer to the question whom the determined performance serves. For example: Do customers of banks benefit from a high return on equity (ROE)? Or is a high ROE a socially desired target? Probably not: As single efficiency scores may imply inappropriate incentives (e.g. indiscriminate cost-cutting, equity reduction etc.; we will discuss this below in section 2.3), we observe against the background of the actual financial crisis that finally society has to bear the burden in the case of banks' failure. So performance measured on the basis of the ROE does not serve society, but rather, for instance, the managers – in the case they receive bonus payments on the basis of ROE.

It is a typical characteristic of the available literature that a description of banks' objectives in the context of the respective research topic is being withheld. Instead, efficiency scores serve as multi-purpose indicators of banks' performance in all aspects, ranging from an assessment of the management's skills to society's interests concerning a sound and stable financial system. We will systematically discuss this mismatch in detail in chapter 3.

For the moment, we will try to shed some light on various and alternative aspects under which banks' performance might be assessed. Banks considerably differ from conventional manufacturing industries in various respects as they fulfil multiple and responsible roles in modern economies. So the basic understanding of what constitutes an 'efficient bank' or even an 'efficient financial system' may differ among a multitude of interest groups. As a result, traditional bank efficiency assessments on the basis of accounting ratios or cost functions are likely not to capture all aspects of banks' responsibilities.

We identified four groups exhibiting a cluster of interests of their own regarding financial institutions[2]. The groups comprise (1) the owners of banks (shareholders in commercial banks, public bodies in savings banks and members of cooperative banks), (2) private and business customers, (3) the banks' managements and (4) society, basically represented by the banking supervision. As a matter of fact, the respective groups have to be understood as stylized entities in the sense of ideal types. As every person participating in business life typically belongs to multiple groups, at least to 'society', we are not talking about the fields of interest of single persons or well-defined groups of people.

This assumption will enable us to exemplarily discuss conflicts of the respective goals of the stylized groups. So we will not only think about possibilities how to measure bank performance in each case – general problems of data availability will be left aside here – but primarily consider why advantages in one field may involve disadvantages for the interests of another group. In other words, we will show that bank efficiency represented by a single figure basically can only reflect unidimensional aspects of performance, although bank efficiency literature typically does not elaborate on this fact.

2.2.1 Interest groups

Shareholders

Owners of banks can basically be divided into two subgroups: Private and public shareholders. Typically, private investors hold shares of certain commercial banks that are listed on the stock exchange. They are generally free in the choice of stocks, thereby pursuing the target of optimizing their portfolio returns. In this way, investments in commercial banks do not substantially differ from investments in other non-monetary firms.

Public shareholders, on the other hand, typically participate in state and savings banks. They cannot readily sell their shares or attempt to increase portfolio value, as public banks are commissioned to uphold public assignments. Generally, they are obliged to pursue the improvement

[2] It is here understood that 'financial institutions' appear as so-called 'universal banks', providing the complete range of services defined in the German *Gesetz über das Kreditwesen* (§ 1 KWG)

of commercial structures in the respective regions even if this affects profitability[3].

Stable returns on their equity investment, i.e. the growth of shareholder value[4], are the basic objective of any shareholder and need no further explanation. Beyond that, we will lay down why we think the assumption of social responsibilities to be another subordinate, but nevertheless important objective of shareholders[5]: While public banks are obliged to meet certain responsibilities anyway, e.g. the provision of financial services to the socially underprivileged, teaching young people how to deal with money etc., commercial banks rather focus on prestigious projects like cultural sponsorship (*Do good and tell people about it*, see, for instance, Deutsche Bank AG (2009)).

The motivation for corporate social responsibility is manifold. And it is not quite clear whether the impetus comes from the owners or from the management. But we can state that for every firm and their owners involved in business especially in the Anglo-Saxon world, charity has a compulsory (Puritan) tradition. Cotton Mather[6] emphasizes a cause-and-effect relationship between social piety and success: *Honor the Lord with thy substance; so shall thy barns be filled with plenty [...] Obscure mechanics and husbandmen have risen to estates, of which once they had not the most distant expectation*[7].

Whether there exists a 'divine dividend' or not, it is not entirely unjustifiable to expect certain positive effects on the business performance. But just as it is the case for all investments in corporate identity, the return cannot be calculated, because the link between social actions and financial performance is not straightforward (Bansal, 2005). At least, what definitely remains for the shareholders is a feeling Andreoni (1990) calls 'Warm Glow'. Certainly, unlike this author resorting to the uncountable microeconomic concept of household's 'utility' which is

[3]For details we refer to the laws of the German federal states (*Sparkassengesetz*).

[4]For definitions of how to measure shareholder value, see Hartmann-Wendels et al. (2007, p. 351 f.).

[5]See Berrone and Gomez-Mejia (2009) for an interesting survey on the different forms of corporate social responsiblities in non-monetary firms.

[6]February 12th, 1663 – February 13th, 1728 in Boston; U.S. Christian leader who was a Puritan (cp. Longman dictionary).

[7]Quoted from: Esmond Wright, Franklin's Philadelphia, London, 1986.

increased by philanthropic activities, we have to stick to more 'profane' values in order to measure banks' performance in this field: e.g. the amount of funds donated, the number of artists, scientists, students, schools subsidized, the foundation volume, the reduction of CO_2 emissions, the working hours in welfare services etc.

Customers

As banks typically have very heterogeneous groups of customers we will limit our examination to the basic groups of private and business customers. Referring to the traditional Intermediation Theory (Sealey and Lindley, 1977), business customers are 'deficit spending units', profitably investing the savings of 'surplus spending units', i.e. private customers.

Surely, the basic need for both groups of customers is the security of their deposits. So normally, we should expect customers to assume the role of a supervisor, compelling the banks to preserve market discipline, i.e. to avoid high risk projects. Among other authors, Macey and Miller (1992) note that in reality customers do not assume this responsibility. The reason is that we have public and private deposit insurance schemes, so that small savers' deposits are supposed to be safe. But unfortunately, the insurance does not only favour risk-taking behaviour of the shareholders, but moreover, it only works when single banks default. Indeed, the actual financial crisis made clear that banks are 'systemic institutions', and so they are likely to default collectively. In this case, nobody knows to what extent the insurance scheme works. Against this background, the security of the deposits is not a specific problem of the group we called 'customers', but concerns the whole society and will be discussed later.

The provision of payment systems is the prerequisite for any commercial transaction. This includes the supply of cash (access to cash terminals), foreign currency exchange, the installation of networked systems of electronic cash and credit cards, booking of debits and transfers, handling of acceptances and checks, as well as the preparation of account statements. In developed economies, all these services are basically assured although associated with different implicit and explicit costs for the customers. Furthermore, there is a certain problem regarding the security of the transactions: As we observe the direct customer contact becoming more and more rare in favour of self-service terminals or internet PCs, new forms of electronic fraud occur (Bundeskriminalamt, 2009, section 4.2.2). So actually it is a primary objective of customers to *safely* conduct banking transactions.

Both private and business customers may get into a situation in which they need additional funds to acquire objects of higher value. Some examples in the case of private customers are consumer credits, car credits or real estate loans. Not only is a low interest rate of interest to the customer but also, in the run-up, a detailed analysis on the side of the bank whether the customer can afford the credit in the long run at all. It is thereby acting as a 'delegated monitor' on behalf of the depositors (Diamond, 1984).

In the case of business customers, the services typically comprise additional funds for investment purposes. If the interest rate is lower than the expected return of the investment the firm will apply for a new loan. So banks can actively determine the investment activity of their customers by setting the interest margin, i.e. the difference between loan and deposit rate. Regularly, the interest margin is subject to competitive pressure. But we might expect customers who remain loyal to a single main bank[8] to pay more interest, because they benefit from competition among banks to a lesser extent.

Customers interested in volatile financial investments need an asset account. The bank provides the corresponding services, i.e. the administration of the shares, bonds, investment funds, certificates etc. for account of a third party. Typically, the bank does not only provide access to trading platforms and calculates the portfolio return but also offers advice and consultancy. Private customers not familiar with portfolio design and available products depend on objective advice on the side of the bank. It is the task of the financial advisor to identify the customer's needs regarding risk attitude and time horizon. In practice, we observe conflicts of interest between the recommendation of optimal products and the recommendation of products with a higher premium for the bank/the advisor. Thus, the resulting portfolio structure may be inappropriate because it misses the customer's objectives.

[8] 'Relationship lending.'

Management

A bank's management is often seen as an intermediary between shareholders and society: On the one hand, managers are obliged to fulfill shareholders' objectives (especially the maximization of value of equity shares), and on the other hand, society restrains their attempts by means of a restrictive regulatory system (especially risk control). In this situation managers have to defend their position on both sides. As a result, we do not only learn from the daily press but also from scientific literature that 'managers are likely to maximize their own utilities by engaging in perquisites consumption and/or other non-optimal expansion of inputs and outputs' (Pi and Timme, 1993). This behaviour might result in less optimal decisions for both shareholders and society if no precautions are taken.

So from the point of view of management, the first objective must be the possibility to do business without interference from outside. The lower the exertion of influence by the political sector or the shareholders, the more opportunities for independent decisions managers will be able to make. Not incidentally, the media do not only report on professional but seemingly amicable relationships between bank managers and politicians. This close relationship is often considered to be inappropriate, if not even suspicious, by the public.

Managers' objectives in the context of our considerations can be reduced to outstanding earnings, i.e. an outstanding basic salary, bonus payments, compensation payments in the case of resignation, and pension claims. They may expect, moreover, non-monetary benefits, like an efficient staff, cars and airplanes at their disposal, a prestigious official residence, media presence, meetings with politicians etc.

Society

The interests of society do not focus on specific banks, but rather on the whole banking system: As we mentioned above, the security of deposits can only be assured against the background of a sound entirety of banks. In fact, we often observe single banks merging on account of financial distress. Normally, the public is not even being informed about problems in the respective banks. And as long as banks solve critical incidents among themselves, the system only proves its stability and reliability. But in the case that multiple banks simultaneously default, the mutual insurance schemes are likely to become insufficient. This is the point where society has to bail out the financial system before the whole economy collapses due to the systemic importance of banks. The resulting banking crisis costs for recapitalization and restructuring can reach about 15-20% of the Gross Domestic Product (GDP) in highly developed countries, and even up to 55% in developing countries (Hoggarth et al., 2001, Ioannidis et al., 2010). There are also additional costs arising from an overall economic slowdown in the subsequent years (unemployment, discontinuity in the flow of credits for investment purposes etc.). Just recently, Haldane et al. (2010, p.87) illustrate: 'The scars from the current crisis seem likely to be felt for a generation.'

But the alternative would be to leave the banks on their own: In this case, panic may lead to bank runs, i.e. the population tries to get their money back from the bank in cash before the bank loses all its assets. The resulting costs for the society are difficult to estimate, as panic might result in chaos. Beyond doubt, real economic problems will arise because even 'healthy' banks can fail during a bank run (Diamond and Dybvig, 1983). So normally the costs for a 'taxpayer bailout' are assumed to be lower than the damage arising from unregulated bank runs.

Against the background of these threatening costs, the stability of the banking system is the main objective of society. Stability is often associated with banks' soundness (Demirgüc-Kunt and Detragiache, 2009). So consequently stability can be quantified by bank-specific soundness scores that can serve as 'early warning indicators' at the same time. Ioannidis et al. (2010) propose and critically discuss several methods how to assess bank soundness[9]. An obvious indicator of financial strength is a rating score from Fitch, Moody's, etc. Whereas the authors still emphasize the accuracy of rating scores, Demirgüc-Kunt and Detragiache (2009) admit that the credibility of such ratings 'has diminished' during the financial crisis. Nevertheless, there do also exist other simple bank-

[9] The authors also argue that the literature on the examination of the systemic banking risk at the country level is still 'problematic'.

specific measures of performance, like the ratio of non-performing loans, the Z-score[10] or the so-called CAMEL-indicators[11].

Surely banks do not only pose a threat to society. On the contrary, banks' services provide the basis for prosperity in the non-monetary sectors by allocating funds for the most profitable use. Moreover, the banking sector itself contributes value to the GDP. Only a few years ago, the measurement of banks' value added posed a serious problem to the system of national accounts. The aim was to identify the value of services the banks provide. As typically banks do not calculate explicit charges on all services, but, to a larger extent, implicitely charge fees by interest premia, an appropriate approach to measure banks' production had to be developed in recent decades: The Financial Intermediation Services Indirectly Measured (FISIM) focuses on the intermediation activities of banks (Eichmann, 2005, Haldane et al., 2010). Accordingly, the value added of a single bank can be calculated as the positive interest margin between loan interest income relative to an opportunity interest income (e.g. measured on the basis of the inter-bank interest rate EURIBOR) plus the positive interest margin between interest actually paid for deposits relative to interest which would have to be paid on the basis of the opportunity interest rate[12]. Recent discussions on the calculation of FISIM expound on the problem of an appropriate opportunity interest rate. In this context, Colangelo and Inklaar (2010) put forward their objection that the low-risk inter-bank interest rates do not reflect the appropriate risk levels banks are actually facing towards their customers. In particular, FISIM is likely to be reported as too high, because 'bearing risk is in general not a productive service as such'. Similarly, Haldane et al. (2010) note that 'it is not clear that bearing risk is, in itself, a productive activity.' The authors show that the current statistical practice of measuring FISIM leads to a surprising outcome[13]: In 2008 we observe the 'paradox of a rapidly rising financial sector contribution to nominal GDP.' The explanation is simple: Due to the economy-wide increase in

[10] The number of standard deviations by which bank returns have to fall in order to wipe out bank equity.

[11] The acronyms of Capital, Asset quality, Management, Earnings and Liquidity.

[12] For a numeric example, see Haldane et al. (2010, p.92).

[13] Cf. Colangelo and Inklaar (2010, figure 6).

the expected level of defaults on loans, banks responded by increasing interest rates. FISIM scores this as a rise in output. Otherwise, a default risk adjusted FISIM adequately captures the pre-crisis ’productivity mirage’ and the resulting peak in 2007 when the risks have materialized.

Finally, it is important to emphasize that society’s interest lies in a *sustainable* form of value added. Unusual growth rates of banking production do not contribute to society’s prosperity in the long run if future generations of taxpayers have to bear the burden of banking system defaults. On this note, the form of value added we are talking about differs fundamentally from shareholder value: Shareholders are able to benefit from occasional increases in value, but society does not.

2.2.2 A ’perverse incentive system’

As taxpayers are held responsible as a last resort for a banking bailout, soundness and stability of the financial system are of particular interest to society. But remarkably, managers’ usual attempts to increase shareholder value might jeopardise exactly this aim. Authors speak of a ’perverse incentive system’ (Macey and Miller, 1992): They refer to the existence of a deposit insurance scheme that fundamentally sets banking firms apart from any other types of firms. On the one hand, deposit insurance prevents bank runs by instilling public confidence. Thus, the public does not feel obliged to control banks. And on the other hand, deposit insurance contributes to excessive risk taking incentives as long as it does not penalize this attitude. Summing up, the authors complain of a lack of market discipline. This induces shareholders ’to use their control position to cause banks to engage in increasingly risky activities in order to transfer wealth from creditors, depositors and the deposit insurance to the shareholders.’ Park and Peristiani (2007) substantiate: It depends on certain conditions that shareholders can turn into the ’enemies’ of regulators. The authors refer to the pioneering work of Robert C. Merton and explain that deposit insurance gives shareholders a put option, i.e. the right to sell the bank’s assets at the face value of its liabilities. Option pricing theory taught us that the value of the put option can be increased by raising a bank’s risk (e.g. low capital ratio, volatile returns; for details, see Freixas and Rochet (2008, p.315)).

So why do shareholders not always induce their managers to engage in risky investments? The reason lies in the existence of a charter value. This is an intangible value that disappears with the closure of the bank and which cannot be recovered in liquidation. Typically, charter value originates when the market value of the assets is higher than the book value of assets. The usual measure for charter value is Tobin's q, i.e. the ratio of the market value of assets to the book value of assets. Empirical results support the moral hazard theory that shareholder value thus takes a convex or even U-shaped form on risk (failure probability of the bank). So if the market value of bank's assets is high shareholders penalize riskier strategies of the management. In consequence, they are the 'allies' of the regulator. If market value decreases the maximization of the put option becomes more interesting and shareholders become the 'enemies' of the regulator at exactly the point when the put option value outweights charter value[14]. So the authors recommend that regulators should intently observe stock price movements to acquire information on the market values of the banks. From Keeley (1990) we learn that against this background especially the degree competition is a crucial factor to destabilize the banking system: Tobin's q can be understood as a kind of monopoly indicator: When the degree of competition is low, market value and shareholder value decrease, and banks have incentives to succumb to the temptation to increase risk, thus threatening the stability of the financial system.

So what do we finally learn from this discussion in the context of our consideration of possible conflicts? First, a high performance in shareholder value is likely to threat society's interest in a sound and stable financial system. Second, the increase of shareholder value does not necessarily conflict with customers' concern to benefit from a high degree of competition. And third, banks in a competitive environment can threaten the stability of the financial system.

2.2.3 The pay-performance sensitivity

We are now describing a central principal-agent problem in banking: Shareholders (principals) basically suspect a management (agents) of

[14]Park and Peristiani (2007) try to empirically identify the respective threshold.

undertaking investment projects that increase their own reputation and income but jeopardize shareholder value. Admittedly, this is a problem which is not specific to banking firms, but deserves special interest in such highly leveraged institutions which are of systemic importance. And as a matter of course, the possibility to shift risk to the principals does in fact exist. The issues that are being discussed in literature – see e.g. John and Qian (2003), Shleifer and Vishny (1997) for a survey – deal with the question whether governmental regulation substitutes or only complements corporate incentive schemes which are meant to harmonize the objectives of managers and owners. So obviously, management compensation is not just a problem between shareholders and their agents, but also concerns society and the regulator, respectively. One can enlarge on the problem mentioned above: Assume that managers have an information advantage over shareholders so that they can hide the true risk of an investment project from the owners and the regulator. Consequently, shareholders are unaware of the actual value of their put option, and the regulator is not able to judge whether shareholders are enemies or allies. Dropping the assumption of a risk-neutral deposit insurance premium John et al. (2000) propose a model which explicitly incorporates a bank's management compensation in the risk-based fair pricing of deposit insurance. As a result, the 'compensation structure is optimal in that it would induce management to undertake Pareto-optimal (value maximizing) investment policies, with no risk shifting.' The compensation structure in turn depends on the level of leverage of the bank.

So additionally to the discussion above, we gained new insights: First, not only the objectives of shareholders, customers and society interact, but also the management's objectives interfere. Second, management is the 'central interface' among the other groups: Assuming that a management's incentive scheme is inadequate shareholders unwittingly engage in high-risk projects, customers do not benefit from increasing competition, and the regulator is left uncertain of the shareholders' attitude. Third, literature provides solutions to harmonize the objectives. And in fact, empirical studies support the hypothesis that shareholder value and management compensation are positively linked, not least because literature further agrees upon the fact that managerial ownership of equity and options in the bank serve to align managerial incentives with shareholder interests. The index being used to assess this question is the so-called pay-performance-sensitivity (PPS). It indicates the dollar increase in a manager's compensation for each 1000 dollars in shareholder value. In most empirical applications, the PPS is positive and ranges from several cents to about ten dollars.

2.2.4 The existence of social dividends

First, we have to clarify the question whether both shareholders' objectives really complement one another. Normally, we should expect the assumption of social responsibilities to put a strain on the growth of bank's equity shares: Berrone and Gomez-Mejia (2009) point at the opinion of Friedman (1970) that any funds that are not invested in productive projects but in non-profit activities reduce a bank's expected payoff. But shareholders know that this view is rather myopic. First, although the exact calculation of a monetary return of philanthropic projects is neither morally desirable nor practically possible, the existence of a return cannot be easily dismissed: Especially in the long run the corporate identity is a crucial factor for any firm to improve its image with customers. Second, shareholders know that the assumption of social responsibilities is the basis for any responsible business in a developed society. Not only against the Anglo-Saxon Puritan background, where philanthropy has a deep-rooted tradition before God (and is expected to be paid back) but also in the European tradition (where philanthropy is an educational task and not expected to be paid back) wealth is inextricably linked with social engagement. Thus, in the German Constitutional Law, paragraph 14, number 2, it says: *Eigentum verpflichtet. Sein Gebrauch soll zugleich dem Wohle der Allgemeinheit dienen.* So after all, and maybe unexpected at first sight, we have to state that both objectives of shareholders do indeed coincide. Even Steve Forbes (publisher of the list of the 400 wealthiest men and women in the U.S.) summarizes: 'In America, business and philanthropy are two sides of the same coin. Commerce means meeting the needs of the people. Philanthropy is the same.'[15]

The assumption of social responsibilities can manifest itself in various ways. Let us discuss the cases in which the objectives of customers and shareholders coincide: At first glance, the assumption of social responsibilities has nothing in common with the basic financial services a bank offers to customers. But we already mentioned that especially savings banks in Germany are legally obliged to provide low-priced or even free banking services to the young and underprivileged. So indeed, a positive interdependency between both objectives exists. In the case of commercial banks we sometimes observe special conditions for students and apprentices, too. But, in our opinion this has nothing do to with social responsiblities, as it is rather some kind of promotion to finally win young customers' loyalty.

The credit availability for customers can be motivated by social responsiblity, too: Recently, there has been a discussion in the media concerning so-called microcredits for start-up entrepreneurs in developing countries[16]. It is about support for the poor who cannot provide any collateral. Normally, the interest rate even for small amounts of money (often less than 100 USD) needed for investment purposes (e.g. grain seed) exceeds expected returns so that the entrepreneurs are finally trapped in debt. Specialized financial institutions, often non-profit organizations, grant loans at better conditions in order to help the poor to help themselves. Although the universal banks we are talking about regularly do not participate in microcredit lending, there does exist the possibility to harmonize the objectives of shareholders and applicants for credit.

Finally, even a bank's management participates in social and cultural sponsorship: It is rather unlikely that the amount of funds donated for charitable purposes has any effect on managers' salaries. On the contrary, Berrone and Gomez-Mejia (2009) describe applied monetary and non-monetary incentive schemes for managers which are meant to support their 'intrinsic motivation' to engage in social and ecological

[15] Quotation found in: Conor O´Clery, Buddy, Can You Spare a Dime? In: Newsweek, March 15th, 2004, p. 14.

[16] Pioneered by Muhammad Yunus from Grameen Bank, Nobel Peace Prize recipient in 2006.

projects. Moreover, making an appearance at sponsored charity events provides a possiblity to easily improve managers' reputation and influence by entering into talks with politicians and the media.

So we may conclude this section: If we (somewhat unusually) decide to measure the bank's performance on the basis of the amount of social benefits, then it is most likely that we also take into account the aspiration not only of shareholder but also of customers and management.

2.2.5 Bank performance assessments: A trade-off

In the preceding sections, we exemplarily discussed several possible conflicts arising among the objectives which different groups might pursue in regard to banks. We limited the description to specific fields of conflict which can be substantiated by references to literature. There can be no doubt that numerous additional conflicts of interest exist. For example:

- Do managers favour competition? Probably not because they know that competition might result in market failure. Or do they rather fear that competition might reduce profit margins and managers' income?

- Do shareholders benefit from sustainable value added? Or do shareholders rather seek for outstanding short-term profits and sell their shares before long-term threats to the soundness of the bank become obvious?

- Does society have to tolerate the outstanding income and political influence of bank managers? Remarkably, bank managers earn more money than can be explained by human capital theories. And moreover, they are additionally rewarded with the possibility to exert pressure on political decision-making processes without being legitimated by a democratic election.

- In how far do customers have to pay higher fees to compensate the bank for the assumption of social responsibilities? This leads back to the question whether charity is another banking output which can be 'sold', or whether shareholders really assume the obligation at their own expense?

As it was to be expected, we found no relevant literature dealing with these issues. So we avoid any further reflections on these topics. But after all, the message has become clear: Any *trustworthy* statements on bank efficiency have to account for either the complexity of competing goals or have to focus on single, well-defined aspects of bank performance.

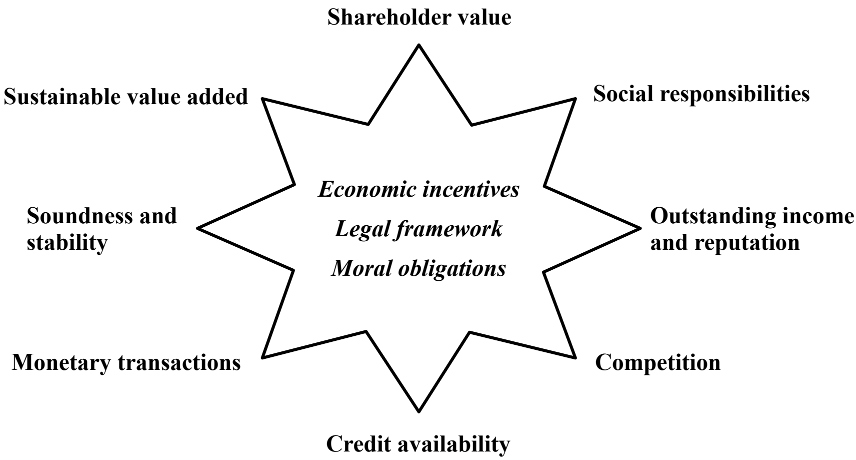

Figure 2.1: Conflicting interests in the assessment of bank performance

So in order to conclude this section, let us step back to get an overview. Figure 2.1 displays some kind of birds-eye view: We arranged all objectives mentioned above in an equal distance to make clear that – without a concrete research topic – there can be no priority. And in fact, we learned that the attempts to 'maximise' the performance in *any* field inevitably affects the other goals and the whole system of competing interests is likely to enter a state of imbalance in the long run. So what keeps the system finally together? In the course of our discussion, we already identified three mechanisms that maintain the objectives in balance.

1. Economic incentive schemes are a subject of Institutional Economics. They support the harmonization of competing objectives by means of market mechanisms, i.e. some kind of self-selection: Behaviour in conformity with the rules will be monetarily rewarded, whereas divergence leads to financial disadvantages. We already discussed this kind of incentive scheme in the context of the moral hazard problem between shareholders and managers. Another application is the assessment of credit risks: Ideally, customers applying for loans are not tempted to conceal their true motives from the bank because this would lead to financial disadvantages in the long run.

2. Not only economic incentive schemes receive legitimation and enforceability on the basis of the legal framework. Wherever market mechanisms, i.e. the incentives to get a reward, cannot be applied, legal norms restrict or prohibit ruthless business activities. We can think of, e.g., a ban on usurious interests, merger control, disclosure requirements, salary limits, etc. In contrast to incentive schemes, compliance with the rules is not being rewarded but violations are being penalized.

3. Beyond the sphere of economy and law participants in the financial markets are obliged to meet basic moral requirements. The achievement of 'moral standards' is rewarded by social appreciation, the under-achievement by social proscription. We already mentioned that the assumption of a certain degree of social responsibility on the part of the bank is self-evident. In contrast, a moral assessment of the core banking activities is rather delicate: In mediaevel times, Christians were even prohibited from granting interest-bearing loans. Today, granting micro-credits to the poor is considered to be charitable. Analogously, it is a matter for debate, for instance, what interest rate is immoral or whether the top managers' salaries are immoral.

So in fact, we disclosed some of the factors Fried et al. (2008) among other authors call 'environmental factors' of performance. The assessment of these factors is surely an issue worth considering in an interdisciplinary context (legal science, social science). But in the subsequent sections (2.3 and also 5.2), we will lay down why these environmental factors have to be strictly separated from the assessment of efficiency. Although both are central determinants of performance, only efficiency is accessible to an econometric approach. We will now have a closer look at the question we have avoided so far: How can we define and classify 'efficiency' in the context of the generic term 'performance' and how can we measure efficiency?

2.3 Measuring efficiency

2.3.1 Common accounting ratios

Traditional accounting ratios are often used by analysts to measure banks' cost efficiency. They are easy to construct and to use, and basically take the form (DeYoung, 1997):

$$\text{Efficiency ratio} = \frac{\text{annual non-interest expenditures}}{\text{annual net revenue}}$$

Annual non-interest expenditures are, e.g., salaries, benefits, materials etc. The annual net revenue is the sum of net interest income ($\equiv$ interest income - interest expenses) plus non-interest income (especially fees[17]). Thus, if a bank A has less expenditures per monetary unit earned than another bank B, the respective efficiency ratio of A is smaller. In other words, and under the assumption of constant returns to scale, bank A is more efficient than bank B[18]. Of course, values ≥ 1 are a cause for concern[19].

---

[17] In recent decades, fee-based activities have increased: E.g. Mutual fund sales, data processing, letters of credit, financial advice, mortgage servicing. Note that alternative accounting ratios of the form $\frac{\text{non-interest expenses}}{\text{total assets}}$ are biased by these activities, because they induce labour costs and add nothing to assets.

[18] Assume again two banks, bank C with the efficiency ratio $\frac{60}{100}$, bank D with $\frac{600}{1000}$. Although the efficiency ratios equal each other, bank D is less efficient if the technology exhibits increasing returns to scale.

[19] Typical values range between 60% and 70% (DeYoung, 1997).

Despite the appealing simplicity of this approach to efficiency, DeYoung (1997), Colwell and Davis (1992), Moormann and Sottocornola (2009) among other authors critizise several aspects of accounting ratios. In particular, DeYoung (1997) admits that on- as well as off-balance-sheet activities are captured, indeed, but he also mentions 'myopic' incentives. He indicates that excessive cost cutting surely ameliorates the efficiency ratio, but on the other hand damages the quality of services. So in the long run business policy oriented towards the minimization of the efficiency ratio will probably go astray.

Moreover, the measure is susceptible to changes of external prices. Normally, we expect efficiency measures to react only to active management decisions (Fried et al., 2008). So if environmental factors exert influence the measure loses its characteristic as an efficiency indicator. DeYoung (1997) notes that a steepening of the yield curve[20] enhances interest margins and drives up net revenue, thereby improving the efficiency ratio. Similarly, Moormann and Sottocornola (2009) point to salaries and wages that are basically determined outside the banks' control. Rising wages affects the non-interest expenditures, worsening the efficiency index. The authors make proposals how to correct the respective efficiency ratios.

Finally, all authors agree upon the need to form peer groups in order to compare the firm-specific efficiency values, i.e. groups of banks operating at the same scale and in the same region, where all banks follow the same strategic objectives and are identically organized. All these environmental factors severely affect revenues and expenses of banks so that comparisons among multiple groups are meaningless. The natural extension of the peer group idea is the regression of accounting ratios conditional on all environmental factors of importance.

2.3.2 Conditional accounting ratio regressions

To circumvent the rather imprecise construction of peer groups within which banks can be compared on the basis of the efficiency ratio DeYoung (1997) suggests a linear regression of

[20]I.e. the term structure of interest rates.

$$\text{Efficiency ratio} = f(\text{assets}, \frac{\text{non-interest income}}{\text{net revenue}})$$

where 'assets' controls for the size of the firm to capture scale effects (and the benefits of so-called 'too big to fail' protections), and the ratio of non-interest income to net revenue controls for the growing importance of fee-based activities biasing traditional accounting ratios (see footnote 17 above). After OLS estimation of the parameters, the expected (mean) efficiency ratio for every bank can be calculated, given the firm-specific values of the respective explanatory variables. A comparison between the conditional mean value and the actual value of the efficiency ratio provides insight whether a bank is better or worse than the average.

In his own application with 330 observations the author achieves a goodness-of-fit R^2 of 19.78%. He admits that a log-specification might lead to a higher explanatory power of the model. Further explanatory variables can be added (even squared variables), but he mentiones that it is important that they are beyond the control of the management. This excludes, for instance, the nonperforming loan ratio.

We found another regression on accounting ratios in Goddard et al. (2004). The authors focus on productivity rather than efficiency. Note that the terms 'productivity' and 'efficiency' are often interchanged in the context of performance measurement. We will see in the next sections that 'efficiency' in scientific literature can also be understood as a target-to-actual comparison.

In the context of accounting ratios, productivity is represented by a ratio of output to input. So the most familiar productivity measure in banking is the Return on Equity ROE. The ROE shares the same shortcomings with the efficiency ratio, i.e. myopic incentives (e.g. reduction of equity, cost-cutting, risky transactions), external price-sensitivity and comparability only within the same peer group.

Nevertheless, as the ROE is the key figure for banks' management, Goddard et al. (2004) discuss a linear model for the determinants of profitability in the form:

$$\text{ROE} = f(\text{ROE last period}, \text{CAR}, \text{OBS ratio}, \text{assets}, \text{ownership dummies})$$

where the new explanatory variables are the capital-to-asset ratio CAR indicating the risk attitude of the bank, the ratio of the nominal value of off-balance-sheet activities to the sum of off-balance-sheet activities plus total assets indicating the extent of fee-based activities, as well as ownership dummies indicating commercial, cooperative and savings banks.

As conditional accounting ratio regressions indeed offer the possibility to compare banks among each other, they certainly provide a rough guidance for practitioners. For scientific purposes, the above-mentioned shortcomings prevail. Moreover, the use of the term 'efficiency' has still remained rather vague so far. We are now going to clarify the basic concepts.

2.3.3 The concept of efficiency frontiers

According to the pioneering work of Farrell (1957), for the first time 'efficiency' was defined within a theory of production based on a microeconomic framework (Kopp, 1981). The author constructed optimal output-isoquants[21] from production data, and defined the radial distance between observed input combination and minimal/optimal input combination as inefficiency, i.e. the waste of resources[22]. Thus, he defined an efficiency measure as:

$$\text{Input based efficiency} = \frac{\text{minimal inputs}|\text{output level}}{\text{observed inputs}|\text{output level}} \in (0,1]$$

As observed inputs can only be equal to or greater than minimal inputs, the efficiency measure ranges between zero and one, with an efficiency score of 100% for the best firms. The optimum isoquant is called a 'frontier function' in the sense of 'undominated performance' (Fried et al., 2008). Figuratively speaking, the frontier constitutes a lower envelope bordering the data cloud. Aigner and Chu (1968) enumerate

[21] I.e. he determined minimal necessary input combinations for a given output level.

[22] To a certain extent this idea was the precursor of the Data Envelopment Analysis DEA with the assumption of constant returns to scale (CRS).

three reasons why firms do not operate on the frontier, but above optimal costs[23]:

1. The existence of random shocks outside the management's control. In the case of the agricultural sector, the authors certainly imply environmental factors such as weather, plagues of insects etc. In the case of financial firms it is worth considering what external factors might be. Surely one may argue that the actual financial crisis is an external shock to the banks. Upon closer inspection it is indeed self-inflicted, although unexpected. We consider 'real' external shocks to be, e.g. wars, revolutions and crime (robbery, fraud). Furthermore, there also exists the possibility that costs are lower than expected: The number of staff on sick leave, for example, may be below average.

   Random shocks surely bias the calculated firm-specific inefficiency scores, because they are not 'real' self-inflicted inefficieny. Insofar, they are unwanted in the course of efficiency estimations. In the case that we have only one observation per firm at hand, a serious problem arises. Only longitudinal observations per firm can give the researcher some indication of single randomly biased observations which do not adequately reflect the management's true ability.

2. The second reason mentioned by Aigner and Chu (1968) is technical inefficiency. The authors hint at input factors technologically no longer viable. In the case of banks the most important input factor is the quality of labour. Generally, the authors assume big firms to have advantages in developing employees' qualifications through extensive training. But we have seen in the course of the financial crisis that even in big banks employees tended to be overstrained by the management of complex financial portfolios. Moreover, also intangible assets such as economic models for the assessment of the risk of credit portfolios can be outdated: We have seen that risk management based on BASEL II specifications basically underestimates the risk of a simultaneous default of multiple creditors. This, together with insufficiently qualified employees, inevitably

[23] Or below optimal production, depending on the context.

results in a poor quality of the outputs, i.e. the banks miss a well-balanced asset structure. Consequently, in order to be insured against unforeseen credit losses (either by equity or by insurance companies) further resources have to be tied up.

3. Economic inefficiency reflects the firms' inability to adjust the output mix to the market situation. Assume that all firms – given a certain market situation – operate on the frontier. After a while the situation might change and some firms constitute a new frontier with lower costs. Now the firms that do not quickly adjust their business plans find themselves with an unfavourably allocated input mix. So what are these 'market situations' the authors vaguely speak of in the case of banks? We can think of increased performance of single input factors, e.g. faster computers and electronic networks, higher educated employees from abroad etc. If the bank is not able to adapt the new input mix soon it will incur higher costs than necessary.

Another interesting question arises: Does the frontier have to belong to the data or is it allowed to run outside the data? Farrell and subsequent authors to date have assumed that the optimum can be delineated by the available data. Consequently, the frontier itself is constituted by observations, implying that the optimum firm(s) can be always found among the data. Although econometricians basically have no choice but to work with the data they have – and can only speculate about the data they do not have – this question is worth considering. In particular, it is about the question whether efficiency should be understood as a normative or a descriptive concept.

All econometric and mathematical methods that constitute an efficient frontier surely imply that efficiency is assessed in a descriptive sense. The fundamental problem that possibly there are no efficient firms in the data sample at all is typically left aside. A reference to asymptotics, i.e. the case of sample size growing to infinity[24], technically solves the problem in such a manner that we have to assume the 'real' efficient firm to be somewhere in infinity. Practically, nobody knows what 'infinity'

[24] See e.g. Wooldrigde (2002, p.7).

means, and, first and foremost, this assumption does not promote the credibility of calculated efficiency scores. So finally, we have to state that efficiency is a relative concept; always relative to the best *observed* firms (Fried et al. (2008, p.11) call this: 'benchmarking the performance of the rest against that of the best').

Even though beyond econometrics, efficiency can also be considered in a normative sense. What performance level are firms supposed to achieve? We have seen in section 2.2 that multiple groups are in a position to impose certain requirements on banks. As long as the expectations can be quantified and the respective firm-specific performance can be measured, the above-given definition by Farrell still holds. Only the data-inherent estimated frontier is replaced by the groups' expectations. A problem arises when the objectives are given as legal norms. For example, German savings banks are subject to the above-mentioned *Sparkassengesetze*. So when it is said, for instance, that savings banks are obliged to support the financial responsibility of young people, the objective cannot be quantified. Eventually, the decision on normative aspects of bank efficiency rests upon moral sense or jurisdiction.

2.3.4 Parametric cost frontiers in banking

Now we take a closer look to the question of how to constitute the efficient frontier (in the case that banks' performance can be measured). Basically, there are two ways to describe the frontier: The parametric and the non-parametric approach. Due to a convenient economic interpretability against the background of microeconomic theory we will exclusively refer to parametric frontiers in the course of this study. Moreover, non-parametric approaches are more or less a topic of the Operations Research while we will strictly focus on econometric methods.

A parametric frontier is given in a functional form. The question arises what variables are on the left-hand side and on the right-hand side, respectively. On the basis of the original approach to efficiency by Farrell (1957), it is an obvious solution to refer to microeconomic concepts of production theory. In particular, Farrell's idea of input based efficiency can be generalized to the concept of cost functions. Cost functions represent the locations of minimum costs given relative prices of the input factors and output quantities[25].

$$\text{costs} = f(\text{input prices, output quantities})$$

In fact, in most applications cost frontiers are estimated. They do not only serve as efficiency standard, but can additionally provide information on the features of the best practice technology as well (Kopp, 1981). As typically all variables are measured in logs, the respective estimated linear coefficients indicate the input price elasticities and the output elasticities. Especially the sum of the output elasticities gives information about the economies of scale in a particular industry (Coelli et al., 2005, p.18 f.).

We found many applications focusing on economies of scale in banking and therefore estimating cost functions, e.g. Allen and Liu (2007), Benston et al. (1982), Berger and Humphrey (1991), Clark and Speaker (1994), Toby (2006). Especially against the background of the actual financial crisis this is an essential question: If there is evidence that big banks benefit from decreasing average costs[26] and privileged access to resources, in a way that they pass the cost advantages on to the customers, it might be the objective of policy makers to support banks' growth. This can be done by tolerantly allowing bank mergers, thereby repressing competition in the financial markets in favour of monopoly structures. Although the authors try to find out the optimal bank size – DeYoung (1997) assumes between 100 and 300 million USD of assets, Clark and Speaker (1994) beyond 1 billion USD of assets – cost functions do not answer the urgent question when banks are too big to fail. In this case, banks are of relevance to the system and will be bailed out by society if they default. Even long before the recent financial crisis, McAllister and McManus (1993) remarked that big banks, in any event, generally follow business strategies that expose them to higher risks.

So obviously, efficiency assessments via cost frontiers serve two purposes: First to establish the cost benchmark in order to measure the 'inefficiency distance' of single firms, and second to derive new insight

[25] Refer to an appropriate textbook on microeconomics, e.g. Beatti and Taylor (1985, p.203 f.).

[26] I.e. reducing costs per unit output by spreading fixed costs over a larger amount of outputs (DeYoung, 1997).

into production technologies from the characteristics of the benchmark that in turn can serve as guidance to policy makers. Now that we have concretized technical characteristics of the frontier, the above-given enumeration of reasons why observations differ from minimal costs can be supplemented by adding technical aspects not considered by Aigner and Chu (1968). Surely, just as random (economic, political) shocks, they are unwanted because they bias the calculated efficiency scores.

4. An obvious reason are measurement errors. Whenever data are collected, mistakes occur. And the researcher has no opportunity to verify the correctness of individual data. Just as in the case of random shocks, only multiple observations (over time) can help to reveal the existence of erroneous data. Moreover, as we will see in the next chapters, stochastic frontier methods exhibit a certain fault tolerance.

5. The cost frontier according to the above-given definition is a function of inputs and outputs of firms. As a matter of course, to appropriately assess the performance of firms, *all* inputs and outputs the firms employ have to be covered by the function. For example, in the case of bottling plants, inputs are labour hours, one or more conveyors, syrup and water. Very easily, we can count the output in litres of lemonade. A firm with more litres of lemonade output than another firm given the same amounts of inputs is obviously more efficient. The situation is different if, in the firm with less output of lemonade, milk is bottled, too. Analogously, in the case of banks, we will learn that exactly the determination of inputs and outputs poses an ongoing problem to literature (Girardone et al., 2004): Whenever firms are benchmarked against the same frontier function, we have to assume that they are using exactly the same inputs and outputs.

The last point mentioned is a very crucial one. Especially in banking we are not only not able to determine a definite set of inputs and outputs, we moreover have to deal with numerous financial institutions that have highly specialized in few products employing very specific inputs. Normally, the researcher is not in a position to subordinate all more or less unique banks in his data set to a single frontier. This would lead to efficiency differences among the banks that are not at all justified. In fact, this discussion has a long history in efficiency estimation – though not specifically in the case of banks. Stigler (1976), in a reply to Leibenstein (1966), vehemently argues that our approach to efficiency resembles a 'tunnel vision of outputs': We impose one person's goal upon firms that have never accepted that goal. Tracing back efficiency changes to employees' motivation and management's ability on the basis of (external) objectives the firms are not even aware of is some kind of a 'shotgun marriage' (Stigler, 1976, p.214). Fried et al. (2008), Førsund et al. (1980) summarize Stigler's statement in the conclusion that *every* measured inefficiency may be a reflection of the analyst's failure to incorporate all relevant variables. In particular, Stigler mentions the concept of 'producing utility'. According to this, a decrease in output may surely be traced back to a change in motivation. But even the wish to avoid unpleasant tasks eventually serves the producer's utility. So at last, there is always an output we can call 'corporate culture'. Stigler's criticism is certainly essential, but has never been picked up in empirical literature. After all, it is not amenable to empirical proof or disproof (Førsund et al., 1980).

Just as we explained in section 2.3.2, facing the unresolvable problem of omitted and immeasurable variables, the authors have recourse to further conditional variables controlling for environmental factors. Surely this does not account for all aspects of banks' specialization and individual product 'preferences', but there is agreement in literature that an appropriate set of control variables justifies a common frontier for a set of banks[27]. So we will complete our list:

6. Banks operating in different competitive environments are subject to a different competitive pressure. On the one hand we expect banks in a lively competition to be in a weak position on the factor

[27] To our knowledge, nobody has ever checked the impact of auxiliary environmental variables on the parameters of a 'pure technical' cost function so far. Cost functions are a microeconomically-founded closed system, whereas 'control variables' are an econometric peculiarity without any theoretical foundation. In particular, we expect the choice of environmental control variables to considerably influence the estimation of cost elasticities and scale economies.

markets (labour, capital), i.e. they have to accept higher costs for inputs. Consequently, they may not reach the minimum frontier costs of banks with high market power. On the other hand, high competitive pressure forces the banks to use inputs very efficiently to stay in competition, because newcomers might capture their market share. Not surprisingly we cannot make any overall assumption concerning the impact of a control variable for the competitive environment (e.g. Herfindahl-Index HHI, Concentration Ratio CR, Lerner Index). Casu and Girardone (2009), for example, found in empirical studies that the degree of competition and efficiency is indeed positively related.

7. The regulatory environment affects firms' possibilities to mix outputs and substitute inputs among each other. So for instance, banks obliged to hold more equity or more deposits at the central bank as reserve than other banks submitted to another regulatory environment cannot reach the same output level with given inputs. In other words, the regulatory framework ties up resources in a non-productive use in favour of systemic stability. Moreover, certain groups of banks might be forbidden to hold only securities without granting loans, although the efficient frontier might be constituted by banks investing only in securities. Even restrictive employment laws belong to the regulatory environment as they can prevent the banks from adjusting/reducing the number of employees if necessary.

8. The last point often mentioned in literature and which affects banks' efficiency is ownership structure. As we already discussed in section 2.2.1, different owners impose individual objectives upon banks. In short, it is reasonable to assume banks obliged to take into account social responsibilities to bear higher costs. It is not surprising, though, that banks exhibiting a totally different ownership structure are rarely benchmarked against the same frontier in literature.

2.3.5 The meaning of 'average frontier'

There remains one big issue that we have not addressed yet. Up to now, we have simply assumed the existence of a frontier constituted by the best firms in the sample. But how are the coefficients of a parametrical frontier determined in practice? Although this is not the place to discuss the estimation techniques in detail[28], we would like to shed some light on the basic procedure and the issues arising.

Basically, the estimation of a cost function linear in parameters as given in section 2.3.4 by means of Ordinary Least Squares (OLS) indicates the average values of costs given output mix and input prices. With reference to Marshall (1920) – mentioned in Stigler (1976) – this characterizes the technology of the representative firm, neither exceptionally successfull nor just struggling. In so far, Marshall argues that firms with fair success and managed with normal ability constitute the production possibility. But literature never followed his proposition. The idea of a Farrell-type frontier does not conceptually coincide with an average function. It is all the more astonishing that even modern frontier analyses are generally based upon average cost functions that are simply shifted to an extreme.

The model usually takes the form:

$$\text{costs} = f(\text{input prices, output quantities}) + \text{inefficiency}$$

where the function $f(...)$ contains the average technological parameters (output and input price elasticities) and the inefficiency-term is a vertical shift-parameter indicating the firm-specific deviation from the unknown frontier. The first simple idea to determine the shift-parameter dates back to Winsten (1957). The author suggested a two-step procedure: First estimating the model by OLS, and second, shifting the regression line downwards so that it passes through the conditionally lowermost observation point of costs. As a result, there is (at least) one firm without inefficiency, i.e. constituting the position of the frontier. The other strictly positive residuals directly indicate the inefficiency of the other firms.

[28] We will turn to this topic in the methodological part.

Together with other studies, e.g. Aigner and Chu (1968), this approach introduced the class of deterministic frontier methods. The authors soon recognized the dependency of their models towards outliers. This attitude arises from the fact that a deterministic frontier basically requires that every detail of the production process is observed (and recorded without mistakes) (Førsund et al., 1980). As this is virtually impossible – we discussed this above – Aigner et al. (1977), Meeusen and van de Broeck (1977) put forward the class of stochastic frontiers (SFA): The assumption that conditional minimal costs are a distribution rather than a fixed value opens the possibility to account for random shocks and omitted variables. Technically, this can be achieved by adding a random variable (usually normally distributed) to the deterministic part $f(...)$ of the cost frontier[29]. This random variable is assumed to be independent of inefficiency, so the estimation of the model requires another assumption regarding the parametrical distribution of inefficiency[30]. Not only Førsund et al. (1980, p.11) note that 'there do not appear to be good a priori arguments for any particular distribution', but we additionally encounter the problem of disentangling random noise and inefficiency for specific firms after the estimation (Jondrow et al., 1982).

Førsund et al. (1980) further substantiate the point that both average cost functions and frontier cost functions reflect one and the same production technology, i.e. that both are conceptually identical. So benchmarking firms against a cost function 'estimated by a novel but complicated method' and not against the simple average does not provide any new insight. In fact, as the authors speak of 'red herring' in this context, we get the impression that parts of the relevant literature consider frontier analyses to be meaningless behind a smoke screen of technical details.

We conclude this part of our preliminary notes with the announcement that we will stick to stochastic frontiers in the course of the next chapters. As they are a standard procedure in literature and excellently developed

[29]Random shocks even allow single observations to lie *below* the cost frontier.

[30]The possibility to estimate a deterministic frontier under the assumption of a specific inefficiency distribution in a one-step Maximum-Likelihood procedure has been discussed earlier in literature, e.g. Afriat (1972). But Greene (1980) showed that this approach is likely to violate statistical properties.

by renowned authors, we cannot ignore this field of application. Although admittedly stochastic frontiers provide certain achievements – e.g. the estimation of efficiency trend parameters (chapter 5) – we will keep the criticisms in mind. In our opinion the application of the newly-discovered Quantile Regression in the course of efficiency assessments overcomes certain drawbacks mentioned above (chapter 6).

2.4 Introduction of the database

2.4.1 Relevance of the Bankscope database

A lot of empirical studies covering bank efficiency estimation are based on balance sheet data provided by the Bankscope database[31], e.g. Girardone et al. (2009), Casu and Molyneux (2003), Pastor and Serrano (2006), Bos and Schmiedel (2003), Lozano-Vivas et al. (2002), Casu and Girardone (2006), Pastor et al. (1997), Weill (2004), Goddard et al. (2004). The publisher claims that the benefit of this comprehensive database is the harmonization of financial accounting standards across the world, as well as the convenient statement in a common currency. This provides a basis to any comparative study assessing the relative efficiency ranking of national bank systems.

Nevertheless, authors working on Bankscope data will soon notice a rather poor data quality: Basically, detailed balance sheet items are hardly maintained. Simple tests for consistency of the remaining data reveal, e.g., that balance sheet items do not add up to total assets or liabilities, the differences of income and expenses do not match profits, or some items are implausibly high or low (or even negative). Consequently, users of Bankscope data implement individual forms of 'data cleansing' before performing their estimation methods. As a matter of fact, the subset of 'clean' data resembles a random sample of observations even across studies covering the same subject. It is no surprise that results show little reproducibility.

Alternatively, authors focusing on specific countries often access high-quality institutional databases provided by central banks. For instance,

[31] Bureau van Dijk, www.bvdep.com.

the Federal Reserve Bank of Chicago freely provides the Reports of Income and Condition (Call Reports) covering numerous banks in the U.S. on the internet[32] (Al-Sharkas et al., 2008, Feng and Serletis, 2009, Kumbhakar and Tsionas, 2008). On the other hand, the German Bundesbank restricts their data to a privileged number of authors, e.g. Koetter and Poghosyan (2009).

In the course of this synopsis, we put forward our approach how to prepare Bankscope data before estimating cost functions for commercial banks in detail[33]. For demonstrative purposes, we exclude savings and cooperative banks from our study. So we circumvent the discussion whether the assumption of a three-pillar banking system – as in Germany – holds for other countries. Moreover, we try to include as many banks in as many countries as possible to exhaust the possibilities in the data set.

2.4.2 Selection of variables

For the sake of example we prepared a dataset of commercial banks in 17 countries with a major occurence of observations in Bankscope: It is Austria (AT), Australia (AU), Canada (CA), Switzerland (CH), Germany (DE), Denmark (DK), Spain (ES), France (FR), Great Britain (GB), Italy (IT), Japan (JP), Luxembourg (LU), the Netherlands (NL), Norway (NO), Russia (RU), Sweden (SE) and the USA (US). We assorted a total of 231 variables, including an identifier, the name, the specialization, the country code etc., as well as all balance sheet items and income and expenses statements listed in the so-called 'Bankscope Global Format'. When we first restrict our observations to hold a positive figure in total assets, we count a number of 3144 banks. The resulting number of observations over the years 2000–2007 add up to 14780 and are distributed as given in table 2.1.

As the number of observations strongly differ from country to country, it is most likely that some countries will be excluded from further analyses when imposing additional restrictions on the data. Moreover, we see that the overall growth of observations over the years is considerably

[32] http://www2.fdic.gov/Call_TFR_Rpts/

[33] Note that this is an illustrating example. Our procedure in the subsequent chapters may slightly differ, according to the respective research objectives.

	2000	2001	2002	2003	2004	2005	2006	2007	SUM
AT	40	46	51	55	65	68	71	62	458
AU	8	8	6	4	8	21	21	18	94
CA	17	13	12	11	11	12	12	4	92
CH	154	155	159	155	162	166	154	132	1237
DE	175	177	173	167	165	170	179	158	1364
DK	45	43	39	39	50	56	58	57	387
ES	23	23	20	15	28	67	59	39	274
FR	150	146	140	136	142	155	137	102	1108
GB	77	76	79	78	111	128	118	83	750
IT	38	36	27	21	93	140	133	100	588
JP	144	143	139	136	137	135	132	128	1094
LU	98	86	83	79	80	79	80	67	652
NL	15	18	20	17	33	31	31	27	192
NO	8	8	8	8	9	9	11	8	69
RU	86	107	114	141	520	721	915	924	3528
SE	5	15	15	14	17	17	18	13	114
US	411	396	374	357	330	324	309	278	2779
SUM	1494	1496	1459	1433	1961	2299	2438	2200	14780

Table 2.1: Observed banks per year and country

biased by the increasing number of observations in Russia. But, as Russia is the only non-OECD country in our dataset, we doubt the economic comparability to the other banking systems. So Russian banks, too, will soon be excluded from our analysis.

Needless to say, only few variables are fully maintained. So the number of banks with a full set of 231 variables will be very small. It is our aim to select the minimum full set of variables which allows us to reconstruct simple balance sheet and income statement schemes and to finally set up common pratice cost functions. In particular, the variables given in table 2.2 are the minium requirement for a reasonably meaningful analysis of bank behaviour.

The table contains the Bankscope description of the variable and the status of maintenance in percent. Wherever applicable, we used variables with an availability of over 90% in our dataset. As it is required all variables to be non-missing, the resulting number of observations is given in table 2.3 on page 42. Astonishingly, only Russian banks still hold a

Variable	**Availability**
Identifier	100.00 %
Year	100.00 %
Country Code	100.00 %
Specialisation	100.00 %
Equity to Total Assets	99.99 %
Return on Average Equity (ROAE)	99.68 %
Cost to Income Ratio	98.65 %
Total Assets th USD	100.00 %
Deposits and Short term funding th USD	99.30 %
Equity th USD	99.99 %
Loans th USD	98.48 %
Other Earning Assets th USD	99.79 %
Total Earning Assets th USD	99.99 %
Non Earning Assets th USD	99.89 %
Other Funding th USD	89.36 %
Non Interest bearing Funding th USD	99.92 %
Net Interest Revenue th USD	99.52 %
Other Operating Income th USD	99.29 %
Profit before Tax th USD	99.66 %
Liquid Assets th USD	99.91 %
Total Risk Assets th USD	99.96 %
Total Fixed Assets th USD	97.50 %
Total Deposits th USD	98.95 %
Total Borrowed Funds th USD	76.87 %
Other Liabilities th USD	99.92 %
Total Liabilities th USD	99.97 %
Total Capital Funds th USD	99.99 %
Total Liabilities and Equity th USD	100.00 %
Interest Income th USD	99.33 %
Total Revenue th USD	99.55 %
Interest Expenses th USD	98.55 %
Personnel Expenses th USD	89.52 %
Other Operating Expenses th USD	81.95 %
Total Expenses th USD	99.53 %
Operating Income th USD	99.57 %

Table 2.2: Observed variables and availability

very high number of observations.

	2000	2001	2002	2003	2004	2005	2006	2007	SUM
AT	17	17	18	14	16	18	19	21	140
AU	2	3	3	1	5	12	17	16	59
CA	4	3				1	1		9
CH	93	77	78	70	69	63	70	66	586
DE	67	70	75	68	77	76	75	67	575
DK	9	3	3	3	11	18	26	31	104
ES	23	23	20	15	22	27	26	23	179
FR		1	1	1	18	28	34	29	112
GB	17	17	19	16	40	46	37	32	224
IT	29	28	18	16	73	114	109	87	474
JP	5	4	5	7	8	8	9	5	51
LU				1	9	14	17	13	54
NL	3	3	3	4	13	13	16	14	69
NO	6	5	8	6	6	6	8	6	51
RU	52	64	75	94	480	693	891	909	3258
SE	4	3	3	2	4	3	3	4	26
US	405	387	365	347	321	315	298	269	2707
SUM	736	708	694	665	1172	1455	1656	1592	8678

Table 2.3: Observed banks per year and country w/o imputation

2.4.3 Data validation and imputation

The reconstruction of some simple balance sheet and income statement schemata offers the possibility of validating a part of the observed values and of imputing some of the missing values. We found the following two schemata for the balance sheet

Assets	Liabilities & Equity
loans	*deposits & short term funding*
other earning assets	*other funding*
non-earning assets	*other non interest bearing*
fixed assets	*equity*
total assets	**total liabilities and equity**

or alternatively

ASSETS	LIABILITIES & EQUITY
liquid assets	*total borrowed funds*
total risk assets	*other liabilities*
fixed assets	*total capital funds*
total assets	**total liabilities and equity**

And for the income statement:

INCOME	EXPENSES
interest income	*interest expenses*
fee + commission income	*fee + commission expenses*
other operating income	*other operating expenses*
	personnel expenses
	loan loss provisions
	other admin expenses
total revenue	**total expenses**

Some simple algebra on the basis of these schemata allows us to unambiguously recover some of the missing values which were curiously 'forgotten' by the Bankscope publishers. Additionally, we are offered the possibility of correcting implausible values. Nevertheless, we found few negative or zero entries in balance sheet items that did not make sense. This is a point where the authors' individual decisions determine the set of banks in the final sample. Usually, the procedure is rather poorly documented, so it is virtually impossible to reproduce any published empirical study. To avoid losing too many observations, we adjusted values < 0 to exactly zero. Table 2.4 shows the remaining number of observations.

Now it is obvious that only few countries can be considered in the course of further analyses. In particular, we chose the six countries with most observations (Russia excluded): CH, DE, ES, GB, IT, US. The resulting dataset comprises a total of 1031 banks.

The application of methods for panel data places high demands on the number of years each bank is observed (individual panel length). Table 2.5 shows how many of the banks are observed in how many years. We see that especially in Spain and Italy none of the banks have records for more than six years, whereas in Switzerland, Germany and the USA

	2000	2001	2002	2003	2004	2005	2006	2007	SUM
AT	19	17	18	15	18	21	23	23	154
AU	3	4	3	3	6	14	19	16	68
CA	4	3	1	1	1	2	2		14
CH	96	86	92	87	87	85	84	76	693
DE	68	75	80	72	80	79	78	69	601
DK	9	3	3	3	12	20	27	35	112
ES	23	23	20	15	24	44	39	30	218
FR	2	3	3	3	20	30	35	30	126
GB	22	21	23	18	46	59	56	43	288
IT	33	29	19	18	82	125	123	96	525
JP	6	6	7	8	12	12	12	6	69
LU				1	12	17	19	17	66
NL	5	6	5	6	22	21	22	20	107
NO	7	6	8	7	8	8	10	7	61
RU	62	81	89	107	476	693	898	910	3316
SE	4	3	3	2	4	3	4	5	28
US	409	391	368	350	324	319	302	273	2736
SUM	772	757	742	716	1234	1552	1753	1656	9182

Table 2.4: Remaining observed banks per year and country

the majority of banks are fully observed. Unfortunately, all banks with only a single or two occurences have to be excluded from our dataset to ensure the full functionality of advanced panel data methods in efficiency estimation. After all, only 793 commercial banks in six countries stand the test of our data cleansing. The respective observations are distributed as given in table 2.6.

Comparing our results to the total number of commercial banks as reported by the Source OECD statistical database[34], we encounter the degree of coverage shown in table 2.7.

2.4.4 Descriptive statistics

Now that we have the full set of observations on the variables given in table 2.2 on page 41, we try to find out something about the banks'

[34]Note that the definitions of what constitutes a 'commercial bank' in both sources do not necessarily have to coincide.

	1	2	3	4	5	6	7	8	SUM
CH	16	17	6	7	10	16	13	45	130
DE	22	22	15	13	12	14	10	28	136
ES	11	12	21	20	8	0	0	0	72
GB	8	16	18	26	5	3	1	5	82
IT	22	43	54	54	3	4	0	0	180
US	23	26	30	31	16	21	31	253	431
SUM	102	136	144	151	54	58	55	331	1031

Table 2.5: Number of banks observed over a period of 1-8 years

	2000	2001	2002	2003	2004	2005	2006	2007	SUM
CH	73	77	89	84	85	83	79	73	643
DE	58	69	72	66	73	70	67	60	535
ES	19	19	20	15	24	28	29	29	183
GB	17	18	20	16	43	49	47	38	248
IT	13	14	15	13	74	104	99	85	417
US	362	365	367	350	324	319	302	272	2661
SUM	542	562	583	544	623	653	623	557	4687

Table 2.6: Observed banks per year and country, final dataset

size, profitability, balance sheet and income statement structure per country and year. All data are given in thousands of US-Dollar; national currencies were converted by means of a single exchange rate to exclude floating effects. Moreover, to avoid the measurement of inflation effects over the years, we have corrected all nominal values by OECD consumer price indices in relation to the base year 2005.

Bank size

First, we set out to learn something about the size of banks in the different countries. For this reason, we defined four classes of bank size between zero and the value of total assets of the biggest bank in the dataset (overall maximum). All class limits are equidistant in logs. The relative share of banks per country and year belonging to each class is shown in figure 2.2. To keep the information clearly arranged, we restrict

	2000	2001	2002	2003	2004	2005	2006	2007
CH	31.60	33.19	39.21	38.71	39.53	38.60	37.80	34.93
DE	28.43	34.67	37.89	37.29	43.45	42.94	41.10	36.59
ES	13.48	13.10	13.99	10.87	17.65	20.14	20.42	19.21
GB	4.16	4.68	5.26	4.49	12.43	14.63	13.99	11.34
IT	4.36	4.49	4.79	4.26	24.50	33.66	31.03	25.91
US	4.31	4.46	4.60	4.46	4.21	4.20	4.04	3.70

Table 2.7: Net coverage final dataset (in percent)

the diagram to the first and the last year in the dataset.

Obviously, the number of small banks decreased between 2000 and 2007. On the other hand, the share of huge and big banks strongly increased, especially in Spain, Great Britain and Italy. In Switzerland, Germany and the USA, the structure of the relative bank size remained more or less unchanged. Unfortunately, as our data represent only a fragmentary subset of the entire number of banks of each country, no further conclusions about bank consolidation and merger activities can be drawn. Some interesting ideas of how to illustrate the distribution of bank size can be found in Ennis (2001). For example, the author refers to 'Gibrat's Law': In the 1930s, Robert Gibrat was the first to formalize a theory of a long-run distribution of firm size. As he assumes the number of firms to be stable over time, and the growth rate of firms to be characterized by an i.i.d. random variable independent of firm size, the distribution of the logarithms of firm size will converge to the normal distribution (convergence of a Random Walk). Ennis (2001) asks whether Gibrat's Law holds for the banking sector. For the sake of simplicity, he avoids the application of extensive tests of Gibrat's assumptions. The author just checks whether the logarithmic size distribution of banks resembles the normal distribution in terms of skewness and kurtosis. We performed the same tests in table 2.8. Obviously, the distribution is positively skewed and has heavier tails. The Jarque-Bera test for normality is rejected every year of our sample. So just like Ennis (2001), we cannot infer any relevance of Gibrat's Law for the banking sector, either.

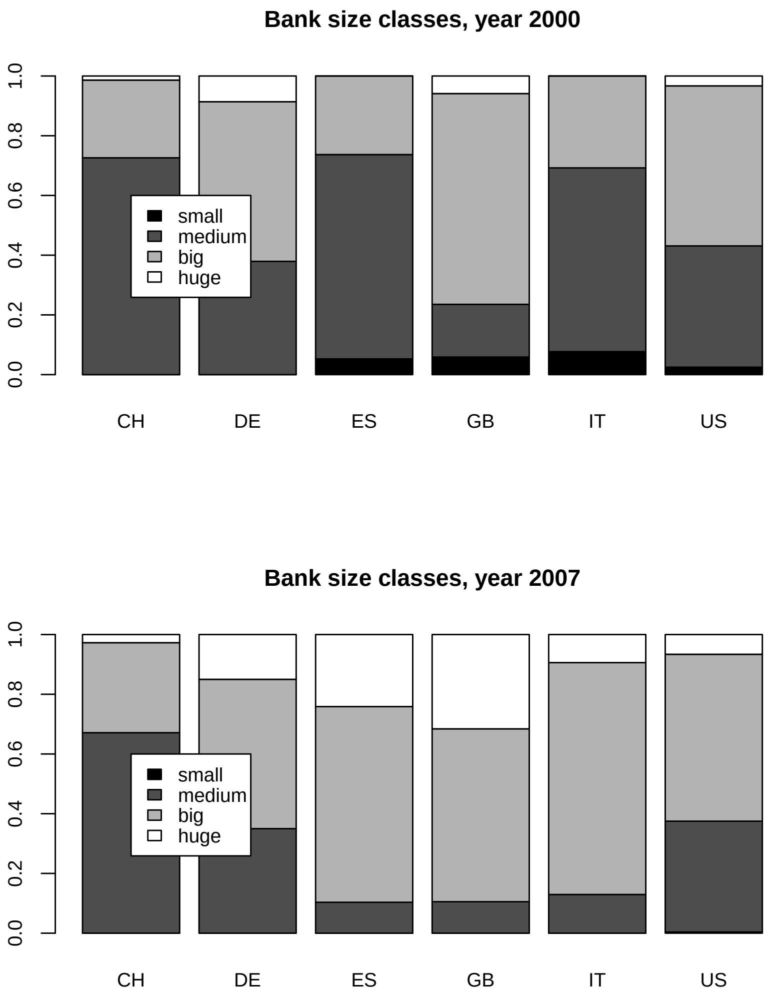

Figure 2.2: Bank size classes

	skewness	kurtosis	mean/median	jbtest
2000	0.1903	3.5958	0.9993	0.0035
2001	0.2892	3.3708	1.0003	0.0040
2002	0.2692	3.3151	0.9978	0.0089
2003	0.3848	3.2173	1.0109	0.0007
2004	0.4074	3.0486	1.0125	0.0002
2005	0.3994	3.1361	1.0108	0.0001
2006	0.2862	3.4142	1.0061	0.0015
2007	0.3782	3.2027	1.0089	0.0008
normal	0.0000	3.0000	1.0000	1.0000

Table 2.8: Test for normality of log total assets

Balance sheet structure

Another interesting point is the structure of the particular balance sheets in an international comparison. To shed some light on the respective weights of different sources of funds on the one hand and the allocation of capital in banking firms on the other hand we turn to the simple balance sheet schemata given above. Summing up all particular items per country and dividing by total assets delivers a stylized mean balance sheet structure for each country. We rejected the consideration of the time structure in favour of a pooled calculation to focus on the differences across countries. Figure 2.3 shows the result.

We basically observe four sources of funding according to Bankscope data availability: Equity, funds at no charge, other funds from the money market, and deposits from customers. Usually, equity is the most expensive source of funding, but also the most persistent. On the other hand, deposits are a bank's favourite as they are usually cheap, but exhibit a high fluctuation. We associate the traditional intermediation business with a high share of deposit funding. In fact, the commercial banks in Switzerland and the USA primarily draw upon deposits as they represent more than 80% of total liabilities and equity. Together with a mean equity share of about 10%, the US-banks seem to display the most conservative funding concept in our sample. On the other hand,

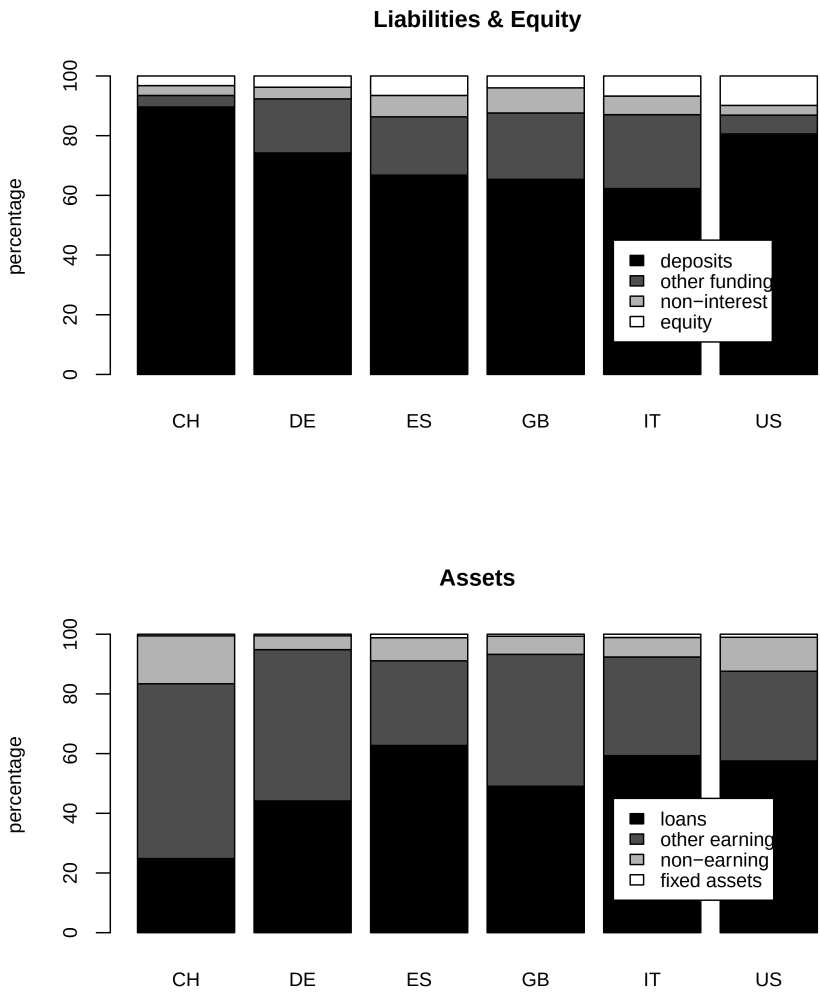

Figure 2.3: Mean balance sheet structure

banks in Italy, Great Britain and Spain lean on money market funding to a greater extent. It is startling to observe that banks in Switzerland, Germany and Great Britain do not even reach a 5% equity coverage of total assets on average.

As for the assets, we expect financial intermediaries to basically provide loans to customers. Actually, only in Spain, Italy and the USA loans represent more than half of total assets. Especially in Switzerland and Germany, we observe banks rather investing in other earning assets, i.e. investment-grade securities, derivatives and insurance assets. The risk structure of money market assets may be much more complex than the credit risk, but the expected surplus yield causes banks to abandon the classical credit business. Not surprisingly, non-earning assets, especially cash and due from banks, should be held in as little quantity as possible to just provide sufficient liquidity. Finally, although always considered to serve as an important input factor in the context of financial production functions, the amount of fixed assets turns out to be rather negligible.

Summing up, our rough guess would be to rate the Swiss and the German banks in our dataset as most vulnerable, while the US banks seem to exhibit a rather conservative balance sheet structure.

Profitability and risk

As the distribution of banks' balance sheet data usually shows a very high variance, the mean balance sheet structures given above might be biased by some single extreme values. To gain some deeper insight into banks' profitability and risk, we calculated the robust median values of Return on Equity (ROE), Cost to Income Ratio[35] (CIR) and Equity Ratio (EQR). The EQR is typically employed as a risk proxy in bank efficiency literature. As is customary, a higher EQR implies less risk, a higher CIR implies higher costs per unit earned and consequently less efficiency[36] and a higher ROE implies higher productivity.

[35]Sometimes called 'efficiency ratio', see above (section 2.3).

[36]Remember that Burger et al. (2008) express doubts that the CIR delivers an authentic reflection of banks' efficiency, as the CIR is – due to its construction – biased by input and output prices outside the management's control, e.g. the price of labour, the interest margin etc.

The respective results per year and country are given in tables 2.9, 2.10 and 2.11. At first sight, it is rather difficult to detect any clear trend over time or any advantages in certain countries. In fact, we observe several breaks that point to data discontinuity rather than to fundamental economic changes. For example, our data show a very volatile trend in ROE for Spain and Italy. In 2003, ROE in Italy is 0.6%, and in the following year, ROE increases to 7.62%. We found no confirmative remarks for this phenomenon in alternative sources. In fact, with reference to table 2.4 on page 44, the reason might be the extremely low number of observations in 2003. Nevertheless, one cannot dismiss the impression that the years 2002 and 2003 are the least productive across all countries. In an overall view, the US banks in our dataset seem to be the most productive, as the ROE remains relatively stable at about 13%, whereas productivity in German banks hardly exceeds 7% in median.

	2000	2001	2002	2003	2004	2005	2006	2007
CH	9.28	6.41	5.17	5.76	5.46	5.65	6.93	6.55
DE	6.38	4.08	3.00	4.12	3.93	6.31	7.37	6.46
ES	4.80	4.29	3.35	5.33	7.11	11.09	11.45	14.69
GB	7.47	7.56	6.83	6.78	11.12	11.86	14.03	10.55
IT	5.00	4.38	2.01	0.60	7.62	8.88	9.85	9.84
US	14.41	13.81	13.95	13.32	13.41	13.56	12.73	10.59

Table 2.9: Return on equity, median values

As the values of the efficiency ratio are somewhat reciprocal to productivity, one should expect high ROE values to be associated with low CIR values. And in fact, in direct comparison tables 2.9 and 2.10 resemble each other like some kind of mirror image. Even the outliers we reported – especially 2003 in Italy – are mirrored. So at least, we can assume our data to be maintained correctly. Our sample might be of low representability in some cases indeed.

The reported values of the equity ratio in table 2.11 are difficult to associate with the tables above. In particular, we cannot confirm the assumption of a negative correlation between ROE and EQR. Whereas in Switzerland, Germany and the USA, EQR values exibit low volatility, we notice a break between 2003 and 2004 in the other countries. Especially

	2000	2001	2002	2003	2004	2005	2006	2007
CH	56.43	61.02	65.86	63.92	62.86	61.70	60.74	60.00
DE	69.15	72.70	71.79	72.79	71.65	73.55	70.43	70.07
ES	68.06	70.46	65.45	61.05	55.11	51.01	47.17	45.34
GB	61.10	57.97	70.21	62.03	54.52	55.82	54.10	57.82
IT	73.08	78.70	83.64	94.02	68.70	63.54	60.85	59.42
US	57.52	58.37	56.54	58.73	58.17	57.23	57.92	60.03

Table 2.10: Cost income ratio, median values

the 10 percentage points loss of EQR in Spain seems rather implausible.

	2000	2001	2002	2003	2004	2005	2006	2007
CH	9.21	8.69	8.71	8.44	8.88	8.40	8.38	8.24
DE	5.04	5.71	5.69	5.79	5.69	5.75	5.38	5.10
ES	13.60	12.53	11.21	18.81	8.22	7.17	6.81	6.46
GB	14.59	15.57	16.89	15.53	8.04	7.69	6.83	6.53
IT	11.48	9.95	13.68	13.29	8.00	7.75	7.33	6.88
US	8.54	8.73	9.00	9.00	9.13	9.37	9.50	9.81

Table 2.11: Equity ratio, median values

To confirm the impressions given above, we calculated the correlation coefficients between all three variables on the basis of the full data, and not just the median values. Table 2.12 shows the result. In fact, the EQR is only negligibly correlated to CIR and ROE. Note that the R-squared of the respective regressions would be only 1%. The correlation between CIR and ROE is marginally higher, as was to be expected.

	EQR	ROE	CIR
EQR	1.00	-0.10	0.12
ROE	-0.10	1.00	-0.45
CIR	0.12	-0.45	1.00

Table 2.12: Correlation matrix: EQR, ROE, CIR

Finally, we learned that our dataset has behaved more regularly over the last years. Any further analyses based on cross sectional data should refer to the years from 2004 to 2007.

Interest margin

After the description of balance sheet structures and commonly used performance ratios we now turn to the discussion of income statement components. Unfortunately, our data prohibit the inspection of detailed income and expense components. So we restrict our analysis to operating results, particularly the interest margin.

According to the idea of financial intermediation banks' primary source of value added is the interest margin between interest paid on deposits and interest received on loans. To give a sense of the evolution of interest margins in our dataset, we calculated the ratio of interest payments to deposit volume per bank. In this way we obtain a proxy for the respective deposit rate. Analogously, the loan interest rate can be approximated by the ratio of interest income to loan volume. The results are given as median values per year and country and are visualized in figure 2.4.

The distance between the upper dashed line and the lower solid line is the positive interest margin. Three characteristics should be commented on: First, all lines exibit a U-shape with a minimum in 2004. As this feature is very striking, we compared our calculated interest rates to the EURIBOR 6-months rate published by the German Bundesbank. Figure 2.5 confirms the minimum interest level in 2004 and 2005. Thus, our dataset seems to contain plausible information on banks' interest income and expenses, reflecting the historical economic situation.

The second feature attracting our attention is the solid line running above the dashed line. This abnormal behaviour occurs in Switzerland (2000, 2005 – 2007) and Italy (2004 – 2007). But as the negative interest margin is not very distinctive, we can also assume slight inaccuracies in the data. It is most likely that in both countries the interest margin was very small. In the case of Switzerland – remember the balance sheet structures given in figure 2.3 on page 49 – low deposit and loan rates become manifest in a large share of deposit financing and little lending activity. This connection does not hold for Italy though. As typically

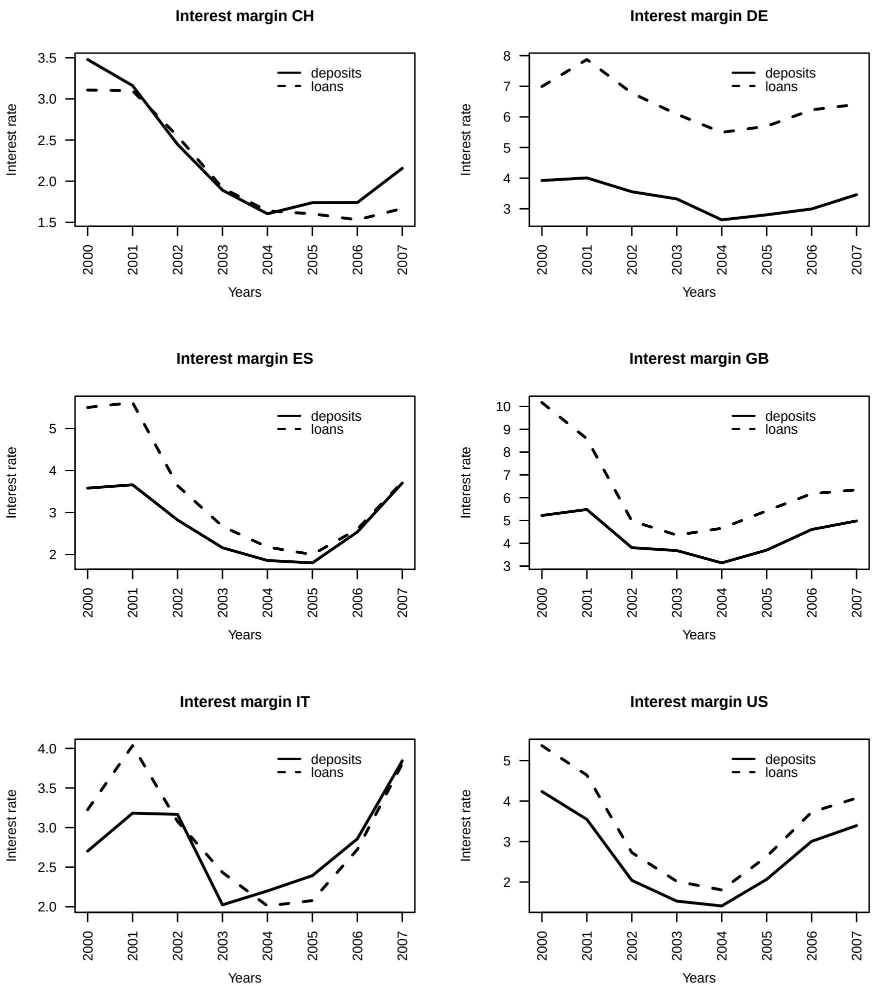

Figure 2.4: Interest margin

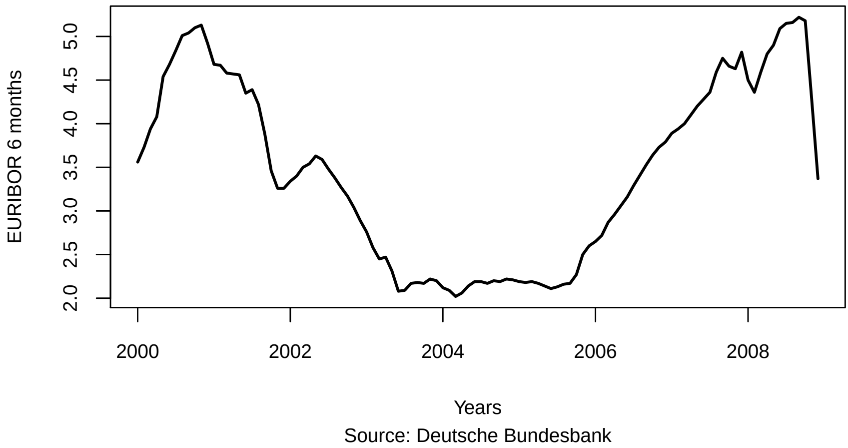

Figure 2.5: Interest level EURIBOR (6 months)

mean interest rates range between 2% and 5%, we observe a very high loan rate surplus in Germany and the UK. Together with a rather low share of loans on the asset side of the balance sheet in figure 2.3, one can guess that lending is limited by the consumers' demand.

The third peculiarity worth mentioning is the absolute interest level in an inter-country comparison. In particular, arbitrage dealers may realize predictable yields by borrowing money in Switzerland or Italy and placing funds in Germany or Great Britain. As we do not expect simple interest arbitrage to be at customers' disposal, we again assume data discontinuity.

2.4.5 Final remarks on data quality

In empirical literature on bank efficiency estimation the presentation of methods and results is usually rated much more highly than the description of the data preparation. Although it is general knowledge in empirical analyses that the treatment of data strongly affects the estimation results, most authors do not give any details on data adjustment nor deletion of observations. Hence, it is virtually impossible for other authors to reproduce the emprirical findings on their own. Finally, the informative value of empirical studies on bank efficiency is questionable.

This short report aims at closing the gap by means of a detailed and comprehensible proposal of how to prepare Bankscope accounting data in order to set up cost functions for efficiency estimation. We showed that the number of usable observations rapidly decreases when imposing certain restrictions on the integrity and plausibility of the data. Our final dataset comprises six countries in the years 2000 to 2007 with 4687 observations on 1031 commercial banks. This is far away from what one may call 'representability'.

We used the resulting sample to learn something about the banking systems in the countries and years we covered. After the presentation of descriptive statistics on banks' size, asset and liability structure, profitability and interest margin in each country, we soon reached the limits of explanatory possibilities. Nevertheless, our data contain enough information to set up commonly applied cost functions for banking firms in the course of the methodical chapters 4 to 6.

3 Chapter 3

Discussion of the empirical assessment of bank efficiency in the frame of a Scheme of Adequation

3.1 Introduction

In the public discussion of the current financial crisis banks are termed 'institutions of systemic relevance'. There is no consistent definition of relevance to the system, though; it apparently goes together with intricate financial links between banks and the other economic sectors. As a consequence this means: if the banks fail, the whole economy fails. In order to prevent the failure of banks policy-maker discuss whether it can thus be legitimate to place on the shoulders of the population and of future generations a burden of additional debts on an unprecedented scale.

The current devices to overcome the financial crisis thus lead to the conclusion that banks belong to the most important economic institutions in the world. In the relevant literature we find a vast amount of studies treating the efficiency of these very institutions. So one might expect a large contribution to the explanation of the current occurences. One must bear in mind that they are the expression of a fundamental failure of the performance of banks. For some decades authors have been aware of the rising importance of banks and the financial interrelations connected with it. On the other hand the results of these studies have never gained public, political or management interest.

In view of the bulk of publications and the lively discussions among renowned authors one cannot presume the research on bank efficiency to be at an early stage of development or its being impaired by fundamental logical flaws. On the contrary: an overview of the wide field of relevant literature provides the reader with a totally consistent picture of bank production in general, of special economic demands on production processes, and of the methods of econometric assessment of efficiency.

The whole literature on bank efficiency, however, seems to thrive in an 'academic bubble' so to speak beside reality. Strangely enough even recent publications do not contain any hints at, explanations for, or perhaps even warnings against the financial crisis. Thus the reader gets the impression that the institutions dealt with in literature are not those banks he reads about in the newspapers. Obviously something about banks is being assessed which does not appear to be real.

How can we ascertain with scientific methods the supposition just put forward? Within the frame of this chapter we would like to enlarge upon the problem and have recourse to 'statistical adequation'. In accordance with Grohmann (1988) we apply a scheme which may prove to be a touchstone of any empirical analysis. By means of this scheme we would like to show if and where in the frame of the analyses the description of banks in a scientific form differs from the image of banks prevailing in politics and economy. In doing so we assume a basic knowledge of the relevant literature on bank efficiency. Alternatively, in order to provide insight into the matter, we refer to the articles available to us by Girardone et al. (2009), Pastor and Serrano (2006) and Fitzpatrick and McQuinn (2005).

It is our aim to provide new impulses to a scientific discussion on the improvement of the meaning of studies on bank efficiency. In view of the current situation of the financial markets we believe research on the activities of banks to be essential as it can provide help to economic policy. On account of the detailed accounting regulations for banks scientists dispose of a rich fund of data especially in central banks and authorities[1]. The methods of efficiency assessment on the basis of contributions of leading econometricians and mathematicians have been implemented in statistical programs in the meantime so that the access to empirical

[1] However also commercial bank data like Bankscope (Bureau van Dijk, Fitch Ratings) are available which are most suitable for the international comparison on account of the consolidation of national accounting standards.

efficiency analyses is open to a wider public. The use of this source ought not to be restricted to the repetition of views perhaps already outdated.

We subdivide this chapter as follows: in section 3.2 we comment on the fundamentals of statistical adequation and the principle of an ideal-type conceptualization in view of the questions to be asked. In section 3.3 we ascertain how to assess the empirical studies on bank efficiency available to us in the context of the schemes laid down. Section 3.4 contains the summary and the outlook.

3.2 Concrete reality and ideal concept

The way of describing objects of research of economic and social analyses shows qualitative differences. First and foremost it is important to differentiate between concrete reality and ideal concepts or, more specifically, to answer the question how appropriately a scientific and concrete description of banks matches an ideal and implicitly underlying concept of banks in the frame of an empirical investigation.

These terms become understandable if one tries to get a general idea of the way social phenomena are characterized: let us assume that in the context of an analysis a scientist sets out to count the number of households in Germany. The scientist is confronted with an unsurmountable problem: on principle 'households' cannot be observed and, thus, are not countable. Although society has an idea of what is characteristic of a household[2] the idea itself cannot be formally laid down. If the scientist, though, operationalizes his endeavor in such a way as to count all German apartments in which according to the residents' registration office a family lives with children, then a definite number can be given. Now it is the task of the scientist who evaluates the data to judge if, to his mind, the value obtained reflects the number of 'households' adequately.

Numerous examples show that this assessment – Menges (1982) talks of a 'semantic recolouring' – is by no means a trivial problem. In those examples no adequate auxiliary quantities of social phenomena are recorded. This corresponds to a dissatisfaction of society which becomes apparent

[2]Thus for instance living in a partnership, with a common budget.

in a fundamental distrust of 'statistical surveys': Brachinger (2007) e.g. quotes the difficulties of an objective assessment of an inflation rate which at the same time serves as a basis for wage negotiations and thus directly determines the purchasing power of the population. Overcoming the 'information deficit observed' [3] between quantities measured and intended objects of research is henceforth to be defined as the criterion of success of the adequation task. A fundamental reasoning has to be undertaken which at first sight does not seem to be related to the literature on bank efficiency. It is the goal of the paper, though, to outline this relevance.

3.2.1 Statistical adequation

A lengthy discussion of the 'statistical adequation' would be out of place in this survey. So we here confine ourselves to a short survey only; for a comprising treatment we refer to Hartwig (1956), Menges (1982), Grohmann (1988) and Litz (1990). Fundamentally, the procedure of adequation describes the task of transforming a real problem into a formal one. Part of this reasoning is not only operationalization in a narrower sense, but first and foremost the decision how the object to be described can be characterized. Or in other words: What is to be assessed. In the reality of social science one uses terms which can be understood as ideal types. What the characteristic features of ideal types are will be expatiated on with reference to the works of Max Weber further down.

Figure 3.1 shows our concept of a comprising adequation 'cycle' in the sense of Grohmann (1988): The operationalization, i.e. the choice of appropriate unities of measurement in order to quantify problems of economic science poses the greatest challenge to the economist. As we shall see later the question in practice is not only which data are desirable but also which data are at all available. With the data at hand the next steps are the choice of an appropriate statistical method and the estimation of an econometric model. As a rule one obtains as a result estimated parameters which in the first place are only figures. With the aid of economic know-how and against the background of operationalization one must succeed in interpreting those figures as a

[3] Brachinger (2007, pp. 10).

solution to the initial question. If no connection can be established with the initial question[4] the operationalization has to be modified and the model has eventually to be re-estimated.

In the course of this paper the scheme described will be applied repeatedly. It is our goal to examine the pros and cons of the existing adequation in assessing bank efficiency as well as alternative possibilities in the scope of this scheme.

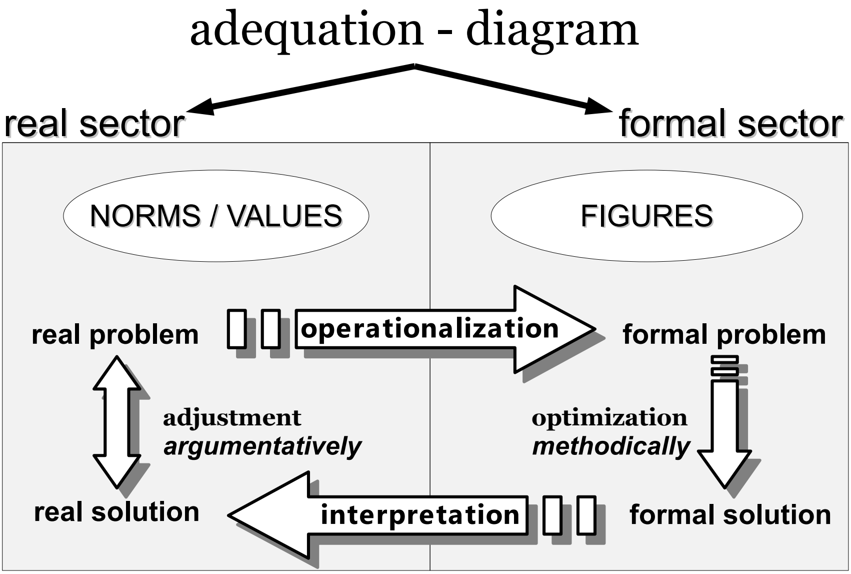

Figure 3.1: Adequation scheme in accordance with Grohmann (1988)

Just one more remark on the term of 'adequation': it goes back to the *idea adaequata* in the sense of De Spinoza[5]. It reflects less the literally 'appropriate idea' rather than the 'perfect idea' in order to characterize the object of investigation. This idea is set apart by its being not only

[4]This need not mean that the figures only contradict the assumption put forward in the description of the problem.

[5]Baruch de Spinoza (1632-1677), dutch philosopher and rationalist.

'true' – more ideally possible than able to be observed – (*idea vera*), but moreover carries ideal-type features which in this purity can never be met with in reality. This very term of ideal type which marks the beginning of every adequation will now be placed into the focus of our attention.

3.2.2 Ideal-type conceptualization

What exactly is characteristic of those ideal-type social terms already alluded to several times?

In modern research one has recourse to the very detailed investigations of Max Weber. In his fundamental article *The objectivity of social science and social policy* dated 1904 Weber[6] lays down the demands on the still recent social and economic sciences. He rejects the expectation empirical sciences can establish 'binding ideals and norms in order to provide advice for practical life' (p. 149) . In other words: it is the goal of social research to provide information free from value-judgements on the 'general cultural meaning of the socioeconomic structure of human social life' (p. 165). This information can be interpreted by the decision-makers themselves and may serve as a basis for political aspirations and changes. Of course, the 'providers' of scientific knowledge are never free from judgements, either. Hence, they ought 'first and at every moment to make very clear to the reader and to themselves which[7] are the criteria applied to reality and from which the value-judgement is derived [...]. Hence, in such cases, the second fundamental condition of scientific impartiality is to let the reader always know that and where the active scientist ends and man with his wishes begins to speak, where arguments are addressed to reason and where they are addressed to the emotions' (pp. 156). This is a sharp distinction between social science as a 'thoughtful ordering of facts' and social policy as 'an exposition of ideals' (p. 157). If scientists neglect the two demands they intermingle science and policy which constitutes 'the most harmful element of our discipline' (p. 157).

So at an early stage Weber is aware of the fact that in the case of uncritical argumentation social sciences may become a mere platform for views and discussions. Therefore he is particulary interested in

[6] The subsequent quotations refer to Weber (1988).

[7] Underlined words are spaced out in the German original.

establishing 'objectivity' in economic and social sciences. This does not mean at all to reject value-judgements in a society: 'The transcendental presupposition of every cultural science lies not in our finding a certain culture or any 'culture' in general to be valuable but rather in the fact that we are cultural beings, endowed with the capacity and a will to take a deliberate attitude toward the world and to lend it significance' (p. 180).

A way out of the dilemma between the identity of cultural being and social scientist is provided by laws as they determine the natural sciences. Weber uses the term as 'a return of certain causal relationships' (p. 171). He is aware, however, of the uselessness of looking for these very connections within the social sciences. The causes and significance of cultural phenomena are not subject to a description of causes and effects: Weber talks of the 'futility of [...] the idea that it might be the goal however remote of cultural science to create a coherent system of concepts which reality might be summarized in and out of which it might then be derived again' (p. 184).

Now we are still faced with the problem of providing the economic and social sciences with an objective *touchstone*[8] if values and laws cannot become the object of investigations. This is the context in which Weber introduces the above-mentioned term of ideal type.

An ideal-type concept is not created as a kind of theoretical average construct of observations of which we dispose. On the contrary, we attempt an abstraction of reality, the achievement of which remains unattainable under the restrictions of society and thus leads to a 'utopia' (p. 190). This is the distinction between ideal type and generic term which endeavours to give us a classification of existing phenomena. No observations can be summarized under an ideal type as it does not contain a general description of features. 'It is a thought construct which is not [...] actual reality and which much less is meant as a scheme in which reality ought to be classified as an illustration' (p. 194). Individual phenomena are selected and undergo a process of abstraction[9]. On

[8] 'As emergency harbours so to speak until one had learned to find one's way on the immense ocean of empirical facts' (p. 206).

[9] 'For everywhere it is the purpose of ideal-type conceptualization clearly to bring out not traits common to elements of reality but on the contrary the characteristics of

closer examination therefore all thought constructs of social sciences are ideal-type concepts.

It has already been mentioned that 'objectivity' implies the absence of personal judgements. 'An ideal type is completely indifferent to evaluation, it has nothing to do with any 'perfection' but a purely logical one' (p. 200). It is this purity which enables us to conceive and describe the essence of things so that a consensus can be achieved which imparts the same concept of the ideal to every observer.

The deliberate accentuation of reality presupposes an imagination which borders on wishful thinking, but at least establishes relationships objectively accessible to reason or, respectively, 'appear to be adequate to our nomological knowledge' (p. 192). Apart from the absence of values there must remain a portion of judgement connected with the ideal type: 'If however we relate this concept (exchange) to the concept of marginal utility for instance, and construct the concept of 'economic exchange' as an economically rational event, this then contains, as every concept does which is logically elaborated, a judgement concerning the 'typical' conditions of exchange' (p. 202). In this process each participant is influenced in his judgement by the society around him, so that the 'idea of modern historically given society based on exchange economy' (pp. 191) must become apparent in the concept of the ideal type. It may assume the aspiration to constitute a model.

In scientific research, ultimatively, the ideal type proves not to be an end in itself[10]. After detailed considerations of the characteristic features of the ideal type we must finally clarify once again its purpose in so far as it has not been discussed in the context of adequation: According to Weber it is the task of research to 'confront empirical reality with the ideal type' (p. 212). A glance at divergencies between ideal type and reality in the aspects mentioned arouses the scientist's interest. So for example an investigation of 'perfect competition' is never of interest, but only the question why in some markets these very conditions do not prevail. The account of deficient competition can only be illustrated with reference to the ideal type. 'All constructions of an essence [...] are ideal

cultural phenomena' (p. 202).

[10] 'The concept of abstract ideal types must be considered not as a goal, but as a means to an end' (p. 193).

types [...] of high heuristic value for research and high systematic value for the construction if they are employed solely as a conceptual means for the comparison and evaluation of reality against their background' (pp. 198).

3.3 Adequation within the empirical assessment of bank efficiency

What purpose do the fundamentals of adequation and ideal type as laid down above serve in the context of an essay on the public acceptance of the assessment of bank efficiency? It is of course apparent at once that the endeavour to assess bank efficiency of whatever kind can only originate from an ideal-tpye classification of what can be conceived of a bank and what its 'efficient' performance may be. Only in view of this ideal-type conception is it possible to investigate into the divergencies why some banks *do not* exhibit the characteristics which we tend to define as efficient bank activities. As long as this ideal-type conception has not been laid down or at least implicitly understood the attempt at assessing efficiency must be regarded as vain. As social values and legal norms conflict with an objective view of the characteristics of banks one reaches here the initial problem of adequation in the context of the empirical assessment of bank efficiency. If the social agreement on the essence of banks fails it becomes impossible to expose the figures obtained (estimated) by the assessment of efficiency as a solution to the real problem.

3.3.1 Banks as a producing business

We now aim at tracing back the attempts at a solution of the task of adequation following the example of the current literature on bank efficiency. As a model we would like to point out the research done by Girardone et al. (2009). We do so not because this research exhibits any particularities but because, on the contrary, in our view and after examination of numerous articles on the subject it constitutes an excellent synthesis of adequation as it is usually done in today's assessment of bank efficiency. Moreover, the article is the most recent study available to us published in a renowned journal. In recent years, the authors have participated in numerous further projects on the subject.

All studies on bank efficiency agree on seeing in banks a kind of 'producing' business with financial products. Thus, whatever the production process may be like the question of input and output factors in this context arises. Girardone et al. (2004) state the problem: 'While the multiproduct nature of the banking firm is widely recognized, there is still no agreement as to the explicit definition and measurement of banks' inputs and outputs'.

Usually there are two well-established solutions describing the production process in banks (Tortosa-Ausina, 2002): the so called Asset Approach according to Sealey and Lindley (1977) and the Value Added Approach according to Berger et al. (1987). In what follows we shall briefly describe either approach having recourse to the findings of Berger and Humphrey (1992)[11].

The Asset Approach according to Sealey/Lindley (1977)

The authors Sealey and Lindley (1977) for the first time set out to delineate a universal concept of banks which they have missed in literature so far[12]. This concept of banks denotes them as institutions aiming at optimizing profits in the sense of production theory. A production process is here characterized as a transformation of input achievements into outputs of higher value. Thus, the component parts of the purely technical outputs (management services, administration, advisory service, trusteeship activities) can be differentiated from the economic output which provides 'genuine' additional value. In the case of banks the process of value added is defined by the borrowing of monetary funds from surplus spending units and the lending of funds to deficit spending units (financial intermediation). If the lending interest rate lies above the

[11] Their statements also contain a reference to the User Cost Approach according to Hancock (1991). We do not know of any empirical application, though, so that a comment on user costs may be discarded here.

[12] Authors like Benston (1965) advocated the view a definition of banks could be adapted to any research project.

borrowing interest rate a value added for the bank is created. Typically, this interest margin is based on a transformation from cheap (usually short-term) money into expensive (usually long-term) money. Porter (1961) calls this 'the very essence of banking, which is to borrow short and lend long'.

Thus the pivotal factor of bank production is the value of the customer deposits. It is characteristic of an input that it causes costs at once but does itself not yet generate yield. This is in fact in accordance with the characteristics of deposits which, on the one hand, must be managed and which, on the other hand, must be lent with interest to customers. One could object that a further characteristic of an input factor should be that in the course of the production process it is deprived of its independent existence. As far as customer deposits are concerned this is of course not the case. Similarly, in order to collect deposits from customers one needs genuine production factors (labour, capital). Hence the authors describe a production process going on in two steps in which customer deposits exhibit the features both of inputs and of outputs as well. In the context of empirical analyses this point is discarded, however, and deposits are given a pure input character.

Berger and Humphrey (1992) note that in this view of banks the aspect of service is completely neglected. Basically, we have to do with accounting and advisory services for the customer. Hence, the Asset Approach they continue is barely appropriate to describe big banks which buy deposits from other banks and important investors in order to grant these funds as credits.

In empirical analyses the Asset Approach (occasionally also called Intermediation Approach) is defined in the following way (Girardone et al., 2009): One can consider as outputs of the production process of banks the value of credits and investment items as listed in the balance sheet. As inputs the authors consider the value of deposits, the value of labour as well as tangible assets. Sometimes, in recent times, one comes across the additional output 'value of off-balance sheet items' (Casu and Girardone, 2005). Hence it is implicitely understood that the study by Sealey and Lindley (1977) published already 30 years ago by means of this modification might be adapted to a modern view of banks. Usually, the above three-input three-output constellation is defined as 'Modified

Intermediation Approach'.

There seems to be a contradiction to the usual conception of production processes in so far as the record kept of production is associated with floating quantities and not with stock values recorded in the balance sheet at a closing date. In this case one might suppose that given an unaltered balance sheet from one year to the next no production has taken place. Still, employees were active who must have achieved a performance. The authors overcome this problem by means of the statement that maintaining the value of the balance sheet item alone can be considered as a performance. Thus one presupposes a steady flow of new investments and allocations of funds in analogy to a bucket full of holes the water level of which would sink if water were not steadily added.

The Value Added Approach according to Berger (1987)

Put in the context of other theories on bank production we find a description of the Value Added Approach in Berger et al. (1987) and Berger and Humphrey (1992). Whereas the Asset Approach or the Intermediation Approach respectively invariably record outputs as assets and inputs as liabilities, there is no such classification with the Value Added Approach: all considerable items in the balance sheet are investigated in view of examining whether the surpluses generated from them exceed the opportunity costs or not (Pastor et al., 1997). In the first case we speak of an output. In principle, it is possible in this context to distinguish between 'important' and 'unimportant' outputs or intermediate products. If the costs exceed the yield we have to do with an input.

In practice the application of this principle is mostly impossible as we seldom dispose of such detailed data from internal bank controlling. For this reason, most authors have recourse to the fundamental results of Berger and Humphrey (1992): Consequently, the two balance sheet items originating from bank production, i.e. deposits and loans exhibit the cost structure of outputs[13]. All kinds of acquired funds (from the financial

[13]In the context of the Intermediation Approach we have already stated that customer deposits also exhibit the characteristics of outputs (an an intermediate product).

markets, from the central bank) are inputs. Moreover, we distinguish the original input factors labour and capital.

Value Added and Asset Approach: a comparison

In view of the empirical studies the difference for the reader between the two approaches lies only in the fact that the item deposits is treated as input in one paper and as output in the other[14]. Must these approaches then be dealt with differently in an ideal-type description of banks?

Surely one may argue that the Value Added Approach is based on another idea of bank activity as no longer the transformation of deposits into loans is the main purpose after all. A problem arises, though, if one tries to uncover the concept of banks as it is alternatively implied by the authors. In the relevant literature there is no hint at an equally reasonable argumentation as that of Sealey and Lindley (1977). In fact, the remarks of Berger and Humphrey (1992) make sense only if one is aware of the reflections on the Asset Approach. Concretely, the uninformed reader would be surprised how one can at all conceive the idea of modelling input and output factors of a production process by means of balance sheet figures.

In the context of our comments on the problem of adequation we can therefore see no advantage in the Value Added Approach itself for the comprehension of bank activities. The pivotal point of the two approaches is the conception that bank activities can be described as the acquisition of 'cheap money' which increases by means of skilful management and 'refinement', so to speak. As all economic production is characterized by refinement the analogy arises to assume in banks, too, a classical production process, which may sufficiently be described by the microeconomic optimization of profits.

Comment on the use of the approaches in literature

Against the background of the two approaches outlined above Tortosa-Ausina (2002) highlights the adequation problem which literature is

[14] Equally, in the context of the User Cost Approach according to Hancock (1991) the decisive question is whether to classify deposits as output or input.

clearly confronted with: hence the problem may be focused around the question which of the two production approaches an author must choose in order to visualize the idea of banks as he conceives it: 'There is no consensus on this point [...] to measure efficiency'. In particular, most authors, as is well known, are concerned with the definition of outputs: Mlima and Hjalmarsson (2002) remark: 'This unresolved question has handicapped the research effort when comparing results from different studies'. Tortosa-Ausina (2002), in fact, no longer restricts himself to the characteristics of deposits, but adds in his final remarks that if only one had more data one would be in a position to measure more products and services. The author is fully aware of the fact that every new definition of production must lead to different results in the context of empirical assessments. He draws the following conclusion from a comparison of efficiency measurements carried out on the basis of the Value Added and the Asset Approach: 'When deposits are also treated as output, savings bank's efficiency is always higher than that of commercial banks [...]. However, the pattern is reversed [...] according to the Asset Approach'. His recommendation runs as follows: 'Consequently, conclusions relative to the efficiency [...] could depend on the model chosen. Our final comments should at least be followed by 'according to our definition of bank output', as our models may not capture the whole range of products and services provided by the banking firm'.

Obviously, the dilemma of the measurement of bank efficiency is well known: We are faced with a complete irregularity of results which can undermine every serious statement on bank efficiency. Astonishingly enough, no conclusion has ever been drawn so far from this state of affairs. At least an attempt might have been undertaken to examine on the basis of some banks which model is adequate to the activities of the banks concerned.

In our view the adequation problem does not lie in the choice of the outputs in the balance sheet of banks. We are not in a position to provide good reasons for deposits to be listed as input or output in a financial production process. The authors in literature as well are obviously faced with a problem which cannot be solved as for decades no consensus has been achieved on a question which appears to be trivial at first sight.

It is worth mentioning that Sealey and Lindley (1977) are the first to criticize the attitude in literature which seems to allow of an ideal picture of banks individually conceived according to one's research project (Benston, 1965). It cannot have been the authors' intention that on the basis of their own proposals new interpretations keep coming up motivated more by the greater availability of data than by serious reflections on the tasks of banks. The reader of studies on bank efficiency needs some practice indeed to perceive how a production process is modelled out of a balance sheet. Incomprehension and general distrust of empirical results arise of course when reading further studies one learns that the choice of balance sheet items can be perfomed in whatever manner.

We suggest to return to Max Werber's ideal-type conceptualization: The approach by Sealey and Lindley (1977) was a well-founded attempt to firmly establish the image of banks prevalent in society. The attempts at modifications by different authors in subsequent years only reveal that the Intermediation Approach no longer appeared to be 'adequate' and that literature tried to adapt itself to a new social image of banks. What was lacking though was to motivate and sufficiently corrobarate the modifications.

3.3.2 Operationalization and estimation

After having expatiated on the starting point of every adequation cycle which consists in the description of the ideal-type conception of what is to be measured we ought to continue with the process of measurement and estimation usually found in literature. These activities are to be classified under 'formal sector' in illustration 3.1.

Concretely, for a judgement on efficiency a production optimum must be identified, e.g. in the manner of a production function. The gap between the under-achievement of firms and their optimal production can then be defined as inefficiency. The usual methods of efficiency measurement on the basis of this concept of efficiency are the parametric Stochastic Frontier Analysis SFA and the non-parametric, non-stochastic Data Envelopment Analysis DEA.

Here again we find numerous aspects which may be dealt with in the context of an evaluation of the validity of bank efficiency studies. On the one hand we have to consider the quality and availability of data (hence the operationalization of an ideal picture by means of appropriate auxiliary 'proxy' variables), on the other hand the choice of a parametric or non-parametric estimation of the respective model.

At this point, however, we should not like to interrupt the purely verbal argumentation by formal aspects. As part of an outlook we would just like to mention that a formal assessment is the motivation for a future paper on specific problems of adequation in the context of efficiency estimation. The reader interested in these questions may refer to Hjalmarsson et al. (1996), Bauer et al. (1998) and Weill (2004) as examples. The authors clearly demonstrate in bank efficiency estimation that with identical data and extreme censorship it is not possible to establish any correlation worth mentioning between efficiency scores derived from different methods. Similar inconsistencies between the results of different methods are reported by Berger et al. (1993).

3.3.3 Interpretation and comparison

We now deal with the last stage of the adequation scheme in figure 3.1. If for every bank concerned we dispose of efficiency values on the basis of the Value Added or the Asset Approach, then these values must be put in relation with the relevant set of problems in the 'real world'. In particular, it must be elucidated whether the values obtained answer the initial question posed by the researcher.

Usually the available data sets are based on some hundreds of banks so that a comment on results of particular banks is unusual. Instead the efficiency scores of certain bank groups of interest are averaged and compared to other groups. Thus typical kinds of questions may run as follows: Which banks are more efficient, public or private ones? Are big banks more efficient than smaller ones? Which (European) country has the most efficient banking system?

As examples we quote from comments on results of some studies available to us:

- *Both parametric approaches suggest that UK credit institutions are more efficient than their Irish counterparts [...] We empirically test the apparent differences in the mean efficiency scores (i) between* *big and small credit institutions and (ii) between UK and Irish credit institutions. A t-test of no significant difference between the two sets of mean efficiency levels is rejected for all models at the 1 per cent level* (Fitzpatrick and McQuinn, 2005).

- *Banks from Germany, Denmark and France (in 1998, 0.659, 0.762 and 0.553 respectively) [...] are the most efficient and banks in Greece, Spain and Italy are [...] among the most inefficient. However, the differences are smaller after discounting the effect of specialisation* (Pastor and Serrano, 2006).

- *Overall, our analysis suggests that differences in cost efficiency across bank types can often be explained by the prevailing financial system in each economy. This evidence illustrates the national diversity of corporate governance systems in Europe and can be important to policy makers who are concerned with the full integration of the European financial system* (Girardone et al., 2009).

These results quoted as examples of empirical bank efficiency studies are indeed answers to the above mentioned typical set of questions. Hence if we wonder for example which financial system was the most efficient in Europe in 1998, we obtain the answer from literature – a litte overdone: Denmark first ($76,2\%$), Germany second ($65,9\%$) etc. These differences in average figures are 'statistically significant' which is apparently meant to exclude that the efficiency scores across countries originate from identical data generating processes. And on the basis of the results obtained policy makers are called upon to choose the financial system – in reality only the totality of banks – of the most efficient country as a model for reform in their own country.

At this point, it is necessary however to raise the question in which way the problem is presented in the studies. Usually institutional changes of political and economic conditions in particular countries or unions of countries (EU) are an occasion to examine the effects on the efficiency of banks. By 'efficiency', in most cases, one implicitly understands the mean efficiency of single institutions as one group, although in a generalizing way one often talks of a 'financial system'. And this efficiency, as stated above, is measured from the point of view of business administration and production theory. Thus, a question initially comprising the economy as a whole is reduced to the production problem of a single firm which may additionally be described as controversial. All other interesting questions as for example concerning the stability of banks can of course not be tackled by the Value Added and Asset Approaches.

Hence between the kind of question and the efficiency values obtained a gap of understanding becomes apparent which must lead to such strange conclusions of authors as quoted above. The deficiency barely lies in the fact if under certain assumptions one can interpret a given figure as the efficiency of a bank. In a particular case, one may even consider it as a plausible characteristic of a bank. But to assume that this idea of production is common to a conglomerate of banks and their common competitive benchmark is more subject to doubt than generally believed.

If in particular cases the authors even talk about investigations into the efficiency of world-wide financial systems, and if in actual fact only production mechanisms of individual institutes are meant, the reader is expected to make a considerable effort when comparing the goal of the investigations with his own expectations. Against this background it is no wonder that the results published do not meet with public interest.

3.4 Conclusion

We should now like to bring to an end our comment on the adequation approach as it is usually employed in literature to measure bank efficiency. In conclusion we have acquired some knowledge of banks which represents them from the point of view of a profit oriented firm in a microeconomic sense. The fundamental idea is that of a production process which can be quantified by means of balance sheet items. A surplus is brought about by the purchase of short-term financial funds which in a different use (long-term loans) enable the bank to realize a profit. Concretely, in literature we come across an operationalization which considers labour and capital as input factors, credits and investment papers as outputs, and deposits inconsistently as input or output.

We are of course faced here with an ideal-type conceptualization of things the realization of which we cannot be expected to observe[15]. We notice as important the social consensus on these assumptions which obviously does not exist. In fact, during the financial crisis banks are perceived by society more as (unsuccessful) merchants of risks on the financial markets and less as producers aiming at profits in the field of credits and securities. In the empirical studies this obvious adequation gap becomes evident in the wish – sometimes somewhat unexpectedly – to take more into account activities beyond the balance sheet.

Here, too, the problem arises that these values as per period end do not allow of a statement about the inherent risk. The grant of a credit or a purchase beyond the balance sheet of a derivative product is assigned a particular nominal value and therewith becomes part of the bank's output of the same accounting period. On this basis the efficiency is measured. If it becomes evident that in the following accounting period this item is without value any more adjustments can be performed. This is done irrespective of the estimation of the preceding period although the wrong decision on the investment was taken then.

Motivated by these deficiencies of the empirical bank efficiency estimation and against the background of the immanent importance of an objective measurement of bank efficiency we have pursued the aim in this study to trigger a debate on the reliability of bank efficiency studies. For this purpose we defined our touchstone, i.e. the principles of an ideal-type conceptualization in the context of an adequation scheme in the first part of our paper. Although the comment on this aspect has led the reader a little away from the goal of bank efficiency estimation we could nevertheless, in the second part of our paper, have recourse to this knowledge in order to faciliate the understanding of the processes.

Thus we can identify the adequation problem in bank efficiency estimation as a fundamental gap of understanding between the claim laid down in recent empirical studies and the expectations of a society eager to obtain an explanation for proceedings observed in reality. The target of many of the studies available to us and destined to compare the efficiency of international financial systems is of course so vague that every reader

[15] We remind the reader of the discussion on the distinction between an ideal type and a generic term.

not acquainted with this branch of literature may develop another idea of 'efficient financial systems'. If in the course of a study it becomes evident that the authors quantify an ideal picture of banks which presupposes an attitude of banks intent on their own individual economic succes and on maximizing profits the reader can be no means recognize the relevance of the results in view of the systemic importance of the entirety of banks. In many cases it may in fact be doubtful whether maximizing short-term profits is economically desirable.

So in accordance with literature, we advocate the view that the classical conception of production processes in banks is outdated given the economic questions we have to raise today. Nevertheless, this discovery may not serve as a licence to occasionally change the operationalization, guided by data availability: The fundamental link between the ideal type formulation of a real economic problem and the adequate measurement of the respective problem must not be interrupted.

Moreover, the very measurement of production on the basis of balance sheet figures at a particular date does not convey the impression of the dynamism of decisions how to allocate funds. Banks are involved today in highly complex financial markets which did not exist when Sealey and Lindley (1977) carried out their research. In this context we would like to draw attention to a paper by Allen and Santomero (1998): The authors identify an ideal of banks as the central risk managers of the national economy. If one pursues this idea the efficiency of a bank can be derived from its ability to generate portfolios with a stable and balanced yield over a period of time. Presumably this approach might meet public interest, although the question of how to operationalize this idea remains a topic for further research.

4 Chapter 4
Methods of cross-sectional stochastic frontier analyses

4.1 Introduction

We will start the methodical part of this survey with a formal description of the basic tools necessary for efficiency estimation: The stochastic frontier analysis SFA is widely used to estimate firm-specific efficiency scores. It is based on the pioneering work of Aigner et al. (1977) and Meeusen and van de Broeck (1977). Kumbhakar and Lovell (2003) provide a comprehensive overview. The fundamental difference to ordinary least squares is the introduction of a two-part error term consisting of a noise and an inefficiency term. For the error as well as the inefficiency term distributional assumptions are made. Most often the half normal assumption is applied, but the exponential, truncated normal and gamma cases are also discussed in the specific literature. While the two-parameter distributions – the truncated normal and the gamma – potentially increase the flexibility of the model, in practical applications, however, problems of identification seem to outweigh the potential gains for either distribution (Greene, 1997, pp. 103), (Ritter and Simar, 1997a,b).

The natural estimation method seems to be Maximum Likelihood (ML) estimation because of the parametric assumptions. But simulation results obtained for the normal-half normal model indicate that a method of moments approach (MOM) (Olson et al., 1980) is superior for small and medium sized samples in combination with inefficiency not strongly dominating noise (Coelli, 1995).

In this chapter we provide detailed simulation results comparing the two estimation methods for both the half-normal and the exponential approach to inefficiency. Furthermore, we compare the sensitivity of the estimation approaches towards misspecification. Our simulations widen those of Coelli (1995) and Olson et al. (1980) for the normal-half normal model in matters of sample size and comprise also the exponential model. The extensive simulation results allow the formulation of rules of thumb for deciding on the estimation approach for normally and exponentially distributed inefficiency terms.

The paper is composed as follows: In section 4.2 we discuss the underlying concept of efficiency and the different approaches to calculate inefficiency scores. In section 4.3 we lay down the results of our extensive simulation studies, especially the suggestions obtained for the choice of the competing estimation approaches in the form of parametrical rules of thumb. Section 4.4 introduces our field of application by first describing the well-established operationalization procedures to obtain efficiency scores for banking institutions in the relevant literature. Next we expose an exemplifying application to a Bankscope dataset for German commercial banks without further expatiating on the preceding discussion of 'Adequation'. Section 4.5 contains the conclusion.

4.2 Efficiency and the stochastic frontier models

In this section we briefly describe the concept of efficiency in the framework of stochastic frontier models based on the half-normal distribution for the inefficiency term. Alternatively, we also specify the exponential model [1].

4.2.1 The concept of output-based efficiency

Farrell (1957) introduced the idea of an empirical approach to relative efficiency by the firm specific quotient of observed production y_i

[1] A closer look at the relevant literature reveals another 'truncated' approach (Bos and Kool, 2006, Battese et al., 2000, Fitzpatrick and McQuinn, 2005): $u_i \sim N^+(\mu, \sigma_u)$ with $\mu \geq 0$. Referring to Greene (1990), Weill (2004) is the only one using a gamma-distributed inefficiency term. But the selection of an adequate distribution of u_i does not have to be overvalued, as Greene (1990) reported extremely high correlations in the efficieny estimates between the half normal, truncated, gamma and exponential models. So obviously there is no need to make use of two-parameter distributions.

to optimal production y_i^*. In conformity with microeconomic theory, production processes are technical relations of employed inputs to maximum attainable output. So when assuming cross sectional data for n units indexed by i $(i = 1, ..., n)$ using K $(k = 1, ..., K)$ different inputs – contained in the input vector x_i – to produce a single output y_i, we can formulate Farrell's idea of technical efficiency:

$$TE_i = \frac{y_i}{y_i^*} = \frac{y_i}{g(x_i;\beta)} \in [0,1]$$

with $g(x_i;\beta)$ as a deterministic production function. It is the aim of the stochastic frontier approach to estimate the underlying technology constituting the production possibilities of a set of firms. We allow a parametric form for the output including stochastic terms

$$y_i = g(x_i;\beta) \cdot e^{v_i} \cdot e^{-u_i}$$

which in logs is

$$\log(y_i) = \log\left(g(x_i;\beta)\right) + v_i - u_i$$

v_i is considered as a normal error $v_i \sim N(\mu_v;\sigma_v^2)$, and $u_i \geq 0$ representing inefficiency.

In the following we assume a simple Cobb-Douglas production function

$$g(x_i;\beta) = e^{\beta_0} \prod_{k=1}^{K} x_{ik}^{\beta_k}$$

which in logs is

$$\log\left[g(x_i;\beta)\right] = \beta_0 + \sum_{k=1}^{K} \beta_k \log(x_{ik})$$

So the output model is given by

$$\log(y_i) = \beta_0 + \sum_{k=1}^{K} \beta_k \log(x_{ik}) + v_i - u_i$$

This leads to firm-specific efficiency scores

$$TE_i = \frac{g(x_i;\beta)\cdot e^{v_i}\cdot e^{-u_i}}{g(x_i;\beta)\cdot e^{v_i}} = e^{-u_i}$$

4.2.2 The normal-half normal model

The component u_i is assumed to be positive representing production inefficiency. Most often u_i is assumed to be half-sided normal

$$u_i \overset{iid}{\sim} N^+(0,\sigma_u^2)$$

The density of u is given as

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

with the moments

$$E(u) = \frac{\sqrt{2}}{\sqrt{\pi}}\sigma_u \qquad \text{and} \qquad V(u) = \left(\frac{\pi-2}{\pi}\right)\sigma_u^2$$

The ML approach for the normal-half normal model

Assuming independence of the error terms v und u the joint density function results as the product of individual density functions

$$f(u,v) = f(u)\cdot f(v) = \frac{2}{\sigma_u\sigma_v 2\pi}\exp\left(-\frac{u^2}{2\sigma_u^2}-\frac{v^2}{2\sigma_v^2}\right)$$

To obtain the density of the composed error term $\varepsilon = v - u$, we first obtain the joint density $f(u,\varepsilon)$. Integration over u results in

$$f(\varepsilon) = \int_0^\infty f(u,\varepsilon)du = \frac{2}{\sigma}\phi\left(\varepsilon\sigma^{-1}\right)\left[1-\Phi\left(\varepsilon\lambda\sigma^{-1}\right)\right]$$

where $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\lambda = \sigma_u/\sigma_v$.

The density distribution of ε is asymmetric and characterized by

$$E(\varepsilon) = E(v-u) = E(-u) = -\frac{\sqrt{2}}{\sqrt{\pi}}\sigma_u$$

The variance of ε is given by

$$V(\varepsilon) = \sigma_\varepsilon^2 = V(u) + V(v) = \left(\frac{\pi-2}{\pi}\right)\sigma_u^2 + \sigma_v^2$$

The log-likelihood finally is

$$\ln L(\varepsilon|\lambda,\sigma^2) = n\ln\left(\frac{\sqrt{2}}{\sqrt{\pi}}\right) + n\ln\left(\frac{1}{\sigma}\right) + \sum_{i=1}^{n}\ln\left[1-\Phi\left(\varepsilon_i\lambda\sigma^{-1}\right)\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\varepsilon_i^2$$

using $\varepsilon_i = \log y_i - \beta\log x_i$.

Having obtained the estimates $\hat{\beta}$, $\hat{\sigma}^2 = \hat{\sigma}_u^2 + \hat{\sigma}_v^2$ and $\hat{\lambda} = \hat{\sigma}_u/\hat{\sigma}_v$, the estimates of the variance components can be recovered:

$$\hat{\sigma}_v^2 = \frac{\hat{\sigma}^2}{1+\hat{\lambda}^2} \quad \text{and} \quad \hat{\sigma}_u^2 = \hat{\sigma}^2 - \frac{\hat{\sigma}^2}{1+\hat{\lambda}^2}$$

The MOM approach to the normal-half normal model

Estimating the production model when using ordinary least squares OLS results in consistent estimates of the slope parameters $\beta_1, ..., \beta_K$, but in a biased estimate of the intercept β_0. As we assume $E(\varepsilon) = 0$ in OLS by definition, the bias is $E(\varepsilon) = -E(u) = -\sigma_u\sqrt{\frac{2}{\pi}}$. Applying the method of moments approach we use the obtained OLS residuals to estimate the central moments m_2 and m_3 in order to adjust the expected bias. The aim is to simply shift the regression line.

$$m_2 = \frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon}_i)^2 \qquad \text{and} \qquad m_3 = \frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon}_i)^3$$

which correspond to the moment equations (Greene, 1997)

$$m_2 = \frac{\pi - 2}{\pi}\sigma_u^2 + \sigma_v^2 \qquad \text{and} \qquad m_3 = \sqrt{\frac{2}{\pi}}\left(1 - \frac{4}{\pi}\right)\sigma_u^3$$

Solving for the variance components results in

$$\sigma_u^2 = \left(\frac{m_3}{\sqrt{\frac{2}{\pi}}\left(1 - \frac{4}{\pi}\right)}\right)^{\frac{2}{3}} \qquad \text{and} \qquad \sigma_v^2 = m_2 - \frac{\pi - 2}{\pi}\sigma_u^2$$

The biased OLS estimate of the intercept $\hat{\beta}_0^{OLS}$ can be adjusted on the basis of the estimate of the standard deviation of the inefficiency term

$$\hat{\beta}_0^{MOM} = \hat{\beta}_0^{OLS} + E(\hat{u}) = \hat{\beta}_0^{OLS} + \hat{\sigma}_u\sqrt{\frac{2}{\pi}}$$

Although the MOM-Estimator is easy to calculate, even without numerical optimization, Olson et al. (1980) note two types of errors occuring when either m_3 is positive (Type I error) or $m_2 \leq ((\pi - 2)/\pi)\sigma_u^2$ (Type II error). A Type I error is likely to occur when σ_u is small ($\lambda \to 0$). This immediately leads to the estimation of a negative variance $\hat{\sigma}_u$ and prevents further calculations. In the latter case, a Type II error does not prevent the estimation of β_0^{MOM}, but causes implausible values of $\hat{\lambda} \to \pm\infty$.

Estimates of individual inefficiencies

As it is impossible to obtain estimates for u_i and v_i simultaneously for each individual firm i, the inefficiency ratio TE_i is obtained as the exponential conditional expectation of $-u$ given the composed error term ε:

$$\widehat{TE}_i = e^{E(-u_i|\varepsilon_i)}$$

The conditional density of u given ε is

$$f(u|\varepsilon) = \frac{f(u,\varepsilon)}{f(\varepsilon)} = \frac{1}{\sigma^*\sqrt{2\pi}}\exp\left(-\frac{(u-\mu^*)^2}{2\sigma^{*2}}\right)\left[1 - \Phi\left(-\frac{\mu^*}{\sigma^*}\right)\right]^{-1}$$

Hence, the distribution of u conditional on ε is $N^{+}(\mu^{*}, \sigma^{*})$, where

$$\mu^{*} = -\frac{\varepsilon\sigma_u^2}{\sigma^2} = -\varepsilon\gamma$$
$$\sigma^{*2} = \frac{\sigma_u^2\sigma_v^2}{\sigma^2} = \sigma^2\frac{\sigma_u^2(\sigma^2-\sigma_u^2)}{\sigma^2\sigma^2} = \sigma^2\gamma(1-\gamma)$$

using $\gamma = \sigma_u^2/\sigma^2$, i.e. the fraction of the variance of the inefficiency to the total variance.

Having obtained the distribution of $u|\varepsilon$, the expected value $E(u|\varepsilon)$ can be used as point estimators for u_i (Jondrow et al., 1982):[2]

$$\hat{u}_i = E(u|\varepsilon) = \left(\frac{\sigma\lambda}{1+\lambda^2}\right)\left(z_i + \frac{\phi\left(z_i\right)}{\Phi\left(z_i\right)}\right)$$
$$z_i = \frac{-\varepsilon_i\lambda}{\sigma}$$

4.2.3 The exponential model

The inefficiency component u_i is assumed to be exponentially distributed. Using the parameterization $\theta = 1/\sigma_u$ the density is given as

$$f(u) = \begin{cases} \frac{1}{\sigma_u}\exp\left(-\frac{u}{\sigma_u}\right) & u \geq 0 \\ 0 & u < 0 \end{cases}$$

The moments are

$$E(u) = \sigma_u \quad \text{and} \quad V(u) = \sigma_u^2$$

[2]Instead of obtaining firm-specific efficiencies from $\exp[-E(u|\varepsilon)]$, Battese and Coelli (1988) propose the alternative estimator:

$$\hat{TE}_i = E(\exp(-u_i)|\varepsilon_i) = \left[\Phi\left(\frac{u_i^*}{\sigma_*} - \sigma_*\right)/\Phi\left(\frac{u_i^*}{\sigma_*}\right)\right]\exp\left(\frac{\sigma_*^2}{2} - u_i*\right)$$

where $u_i^* = -(\log y_i - x\prime\beta)\sigma_u^2/\sigma^2$ and $\sigma_*^2 = \sigma_v^2\sigma_u^2/\sigma^2$. Note that in general $\exp[-E(u|\varepsilon)] \neq E(\exp(-u_i)|\varepsilon_i)$. Furthermore, both estimators are unbiased, but inconsistent estimators because $Var(\hat{u}_i) \neq 0$ for $N \to \infty$.

The ML approach for the normal-exponential model

Assuming again the independence of the error terms v und u the joint density simply results in the product of the two density functions

$$f(u,v) = f(u)f(v) = \frac{2}{\sigma_u \sigma_v 2\pi} \exp\left(-\frac{u}{\sigma_u} - \frac{v^2}{2\sigma_v^2}\right)$$

To obtain the density of the composed error term $\varepsilon = v - u$, we first obtain the joint density $f(u,\varepsilon)$ and integrate out u from the joint density

$$f(\varepsilon) = \int_0^\infty f(u,\varepsilon)du = \frac{1}{\sigma_u}\Phi\left(-\frac{\varepsilon}{\sigma_v} - \frac{\sigma_v}{\sigma_u}\right)\exp\left(\frac{\varepsilon}{\sigma_u} + \frac{1}{2}\frac{\sigma_v^2}{\sigma_u^2}\right)$$

The density distribution of ε is asymmetric and characterized by

$$E(\varepsilon) = E(v-u) = E(-u) = -\sigma_u$$

The variance of ε is given by

$$V(\varepsilon) = \sigma_\varepsilon^2 = V(u) + V(v) = \sigma_u^2 + \sigma_v^2$$

Assuming independence across subjects i, the likelihood is the product of the individual densities $f(\varepsilon)$:

$$L(\log y|\sigma_u^2, \sigma_v^2) = \frac{1}{\sigma_u^n}\exp\left(\frac{1}{2}\frac{\sigma_v^2}{\sigma_u^2}\right)^n \prod_{i=1}^n \left[\Phi\left(-\frac{\varepsilon_i}{\sigma_v} - \frac{\sigma_v}{\sigma_u}\right)\exp\left(\frac{\varepsilon_i}{\sigma_u}\right)\right]$$

And the log-likelihood is

$$\begin{aligned}\ln L(\log y|\beta, \sigma_u^2, \sigma_v^2) &= -n\log(\sigma_u) + n\frac{1}{2}\frac{\sigma_v^2}{\sigma_u^2} \\ &+ \sum_{i=1}^n \left[\log\Phi\left(-\frac{\log y_i - \log x_i'\beta}{\sigma_v} - \frac{\sigma_v}{\sigma_u}\right) + \frac{\log y_i - \log x_i'\beta}{\sigma_u}\right]\end{aligned}$$

The MOM approach to the normal-exponential model

Just as in the case of the method of moments approach discussed above, OLS residuals are used to estimate the central moments m_2 and m_3, which correspond to (Greene, 1997)

$$m_2 = \sigma_u^2 + \sigma_v^2 \qquad \text{and} \qquad m_3 = -2\sigma_u^3$$

Solving for the variance components results in

$$\sigma_u^2 = \left(-\frac{m_3}{2}\right)^{\frac{2}{3}} \qquad \text{and} \qquad \sigma_v^2 = m_2 - \sigma_u^2$$

Analogous to the half-normal case, a Type I error occurs when $m_3 < 0$, and a Type II error when $m_2 < \sigma_u^2$ (virtually impossible). The biased OLS estimate of the intercept $\hat{\beta}_0^{OLS}$ can be adjusted using the estimate of the standard deviation (equal to the mean value) of the inefficiency term

$$\hat{\beta}_0^{MOM} = \hat{\beta}_0^{OLS} + \hat{\sigma}_u$$

Estimates of individual inefficiencies

As the conditional distribution $f(u|\varepsilon)$ is $N^+(\tilde{\mu}, \sigma_v^2)$ and given by

$$f(u|\varepsilon) = \frac{f(u,\varepsilon)}{f(\varepsilon)} = \frac{\exp[-(u-\tilde{\mu}^2)/2\sigma^2]}{\sqrt{2\pi}\sigma_v \Phi(-\tilde{\mu}/\sigma_v)}$$

with

$$\tilde{\mu} = -\varepsilon - \frac{\sigma_v^2}{\sigma_u}$$

the expected value of inefficiency u – given the residual ε in the normal-exponential model – can be written as (Kumbhakar and Lovell, 2003):

$$E\left[u_i|\varepsilon_i\right] = \tilde{\mu}_i + \sigma_v \left[\frac{\phi(-\tilde{\mu}_i/\sigma_v)}{\Phi(\tilde{\mu}_i/\sigma_v)}\right]$$

4.3 Simulation

We apply the two estimation approaches outlined above to obtain estimates of individual efficiencies $\widehat{TE}_i = \exp[E(-u_i|\varepsilon_i)]$. To assess the performance of the efficiency score estimation we calculate the mean square error and the mean average error

$$mse = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{TE}_i - TE_i\right)^2 \qquad\qquad mae = \frac{1}{n}\sum_{i=1}^{n}|\widehat{TE}_i - TE_i|$$

using true TE_i and estimated efficiency scores $\widehat{TE}_i$.

Our simulation study allows to assess the relative performance of the ML and MOM estimators for different n. Additionally, we analyse the effect of λ, i.e. the relation of inefficiency variance to variance of the normal noise, on the appropriate choice of the estimation method.

4.3.1 Simulation design

Now we analyze the estimation of individual inefficiency scores by means of Monte Carlo simulations based on $m = 2000$ replications using a standard simulation setting:

$$y_i = 1 + x_1 + x_2 + v_i - u_i$$

with

$$v_i \sim N(0, \sigma_v) \quad \text{and} \quad u_i \sim N(0, \sigma_u) \quad \text{or} \quad u_i \sim Exp(\sigma_u)$$

$$\sigma_u = \{0.283, 0.447, 0.526, 0.566, 0.587, 0.600, 0.614, 0.620\}$$

$$\sigma_v = \sqrt{0.4 - \sigma_u^2} \quad \text{and} \quad \lambda = (0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0)$$

The inputs x_1, x_2 are drawn independently from a uniform distribution on the interval $(0, 1)$. Sample sizes are

$$n = \{25, 50, 75, 100, 150, 250, 500, 1000\}$$

The computations are performed by means of the **R** Environment (R Development Core Team, 2010).

To assess the robustness of the half normal and the exponential models towards misspecification, we add two misspecification scenarios. In scenario M1 we (falsly) apply the half normal model on data generated under the exponential assumption. Conversely in scenario M2 we (falsly) estimate the exponential model despite inefficiency terms u_i in fact drawn from the half normal distribution.

4.3.2 Simulation results in comparison

The normal-half normal case

n	25	50	75	100	150	250	500	1000
λ	method of moments							
0.5	0.1801	0.1742	0.1727	0.1686	0.1637	0.1568	0.1503	0.1426
1.0	0.1964	0.1877	0.1780	0.1758	0.1685	0.1590	0.1459	0.1348
1.5	0.1879	0.1697	0.1584	0.1521	0.1390	0.1298	0.1214	0.1183
2.0	0.1752	0.1499	0.1368	0.1267	0.1188	0.1113	0.1073	0.1054
2.5	0.1649	0.1308	0.1184	0.1092	0.1028	0.0981	0.0952	0.0935
3.0	0.1531	0.1185	0.1049	0.0991	0.0931	0.0884	0.0853	0.0838
4.0	0.1375	0.1028	0.0896	0.0841	0.0787	0.0741	0.0708	0.0686
5.0	0.1253	0.0938	0.0808	0.0751	0.0697	0.0643	0.0604	0.0582
λ	maximum likelihood							
0.5	0.2440	0.2114	0.1962	0.1883	0.1774	0.1627	0.1496	0.1362
1.0	0.2460	0.2123	0.1935	0.1883	0.1743	0.1584	0.1442	0.1330
1.5	0.2172	0.1809	0.1645	0.1553	0.1395	0.1294	0.1212	0.1181
2.0	0.1898	0.1535	0.1374	0.1259	0.1177	0.1101	0.1067	0.1052
2.5	0.1724	0.1299	0.1152	0.1057	0.1000	0.0961	0.0944	0.0932
3.0	0.1523	0.1141	0.0997	0.0939	0.0891	0.0857	0.0838	0.0833
4.0	0.1290	0.0925	0.0804	0.0762	0.0726	0.0701	0.0687	0.0679
5.0	0.1118	0.0788	0.0693	0.0653	0.0618	0.0598	0.0579	0.0576
λ	advantage							
0.5	MOM	MOM	MOM	MOM	MOM	MOM	MLE	MLE
1.0	MOM	MOM	MOM	MOM	MOM	MLE	MLE	MLE
1.5	MOM	MOM	MOM	MOM	MOM	MLE	MLE	MLE
2.0	MOM	MOM	MOM	MLE	MLE	MLE	MLE	MLE
2.5	MOM	MLE	MLE	MLE	MLE	MLE	MLE	MLE
3.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
4.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
5.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE

Table 4.1: Mean Average Error normal-halfnormal approach

To assess the accuracy of the estimation of the inefficiency terms, we calculate mean absolute differences between estimated and true efficiency scores (*mae*, corresponding mean squared deviations *mse* are reported in the appendix.) for all sample sizes n and λ. Table 4.1 shows the results subdivided into three blocks: The first block gives the *mae* in the MOM-case, the second block in the ML-case and the third block indicates the superior estimation method in terms of a smaller error. Our findings confirm the results of Coelli (1995) and Olson et al. (1980). Due to improved computer capacity in the last decade, we were in a position to perform more extensive computations. Based on these extended simulation results including larger sample sizes and more simulation runs, we confirm their main findings: MOM-estimation is found strongly superior for rather small n and small λ. This is in conformity with intuition, because a small λ implies negligible inefficiency σ_u in comparison with dominating σ_v, which comes close to the classical OLS assumption. In this case, OLS provides the best linear unbiased estimators for $\beta_1, ..., \beta_k$, while β_0 is just slightly biased. ML-estimation should be preferred for larger n and larger λ. But a look at the simulation results reveals only small differences in performance for larger n, rendering the choice of estimation methods rather unimportant.

In general, the mean average error between estimated and true efficiency scores decreases with increasing sample sizes as well as with increasing λ. For small sample sizes $n = 25$ and $\lambda = 0.5$ we find a mean average deviation about three times the size compared to the case $n = 1000$ and $\lambda = 5.0$ for the MOM approach. In case of ML-estimation, which is found considerably inferior for small n, mean average deviations for small n, λ-combinations are about four times the value obtained for large n, λ.

The normal-exponential case

The results of the normal-exponential model are illustrated in table 4.2. The findings of the normal-half normal and the normal-exponential model resemble each other. Again small sample sizes and a small variance ratio λ strongly suggest application of MOM-estimation. But obviously, we observe more n, λ-combinations for which ML-estimation is superior.

n	25	50	75	100	150	250	500	1000
λ	method of moments							
0.5	0.1630	0.1598	0.1566	0.1535	0.1509	0.1443	0.1378	0.1324
1.0	0.1806	0.1622	0.1503	0.1453	0.1412	0.1370	0.1349	0.1338
1.5	0.1669	0.1422	0.1341	0.1291	0.1250	0.1217	0.1195	0.1180
2.0	0.1551	0.1305	0.1209	0.1169	0.1125	0.1086	0.1053	0.1032
2.5	0.1489	0.1197	0.1118	0.1075	0.1022	0.0977	0.0938	0.0913
3.0	0.1419	0.1137	0.1051	0.1002	0.0950	0.0894	0.0850	0.0820
4.0	0.1355	0.1069	0.0972	0.0909	0.0843	0.0779	0.0722	0.0683
5.0	0.1355	0.1023	0.0930	0.0857	0.0781	0.0709	0.0640	0.0594
λ	maximum likelihood							
0.5	0.1888	0.1729	0.1655	0.1616	0.1557	0.1470	0.1385	0.1326
1.0	0.2006	0.1695	0.1539	0.1471	0.1410	0.1366	0.1343	0.1336
1.5	0.1729	0.1408	0.1303	0.1246	0.1210	0.1188	0.1175	0.1170
2.0	0.1534	0.1211	0.1111	0.1078	0.1047	0.1030	0.1017	0.1012
2.5	0.1376	0.1057	0.0976	0.0945	0.0920	0.0901	0.0888	0.0883
3.0	0.1234	0.0948	0.0878	0.0845	0.0820	0.0800	0.0785	0.0780
4.0	0.1076	0.0793	0.0728	0.0698	0.0666	0.0648	0.0636	0.0630
5.0	0.0980	0.0694	0.0630	0.0598	0.0566	0.0547	0.0534	0.0527
λ	advantage							
0.5	MOM	MOM	MOM	MOM	MOM	MOM	MOM	MOM
1.0	MOM	MOM	MOM	MOM	MLE	MLE	MLE	MLE
1.5	MOM	MLE	MLE	MLE	MLE	MLE	MLE	MLE
2.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
2.5	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
3.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
4.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
5.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE

Table 4.2: Mean Average Error normal-exponential approach

Again, the mean average deviation between estimated and true efficiency scores decreases with increasing sample sizes as well as with increasing λ. Similar results were obtained for the half normal model. We also find for small sample sizes $n = 25$ and $\lambda = 0.5$ a mean average deviation about three times the size compared to $n = 1000$ and $\lambda = 5.0$. Just as in the normal-half normal case MOM strongly outperforms ML-estimation for very small λ, while the preferability of ML is based on very small performance differences only.

Misspecification scenarios

As it is an impossible task to determine any real inefficiency distribution across an industry, we can basically assume an underlying misspecification in every applied Stochastic Frontier Analysis. To exemplify the impacts of misspecification against the background of MOM vs. ML-estimation, we interchanged the data generating processes of the normal-exponential and normal-halfnormal case. So table 4.9 given in the appendix shows the *mae* of an exponential model estimated as halfnormal (misspecification scenario $M1$), and table 4.10 in the appendix contains the results of a halfnormal model estimated as exponential (misspecification $M2$).

The findings are straightforward: The exponential model $M1$ shows the predominant advantage of ML-estimation, even more clearly than in the correctly specified case. Obviously, the *mae* of MOM-estimation improves with increasing n or λ, but does not decrease in jointly larger n, λ-combinations. The indications in the halfnormal model $M2$ suggest slight advantages of MOM-estimation in the face of an overall lower error.

Rules of thumb

To summarize the particular advantages of ML- or MOM-estimation, we estimated multiple linear regressions based on tables 4.1 and 4.2 for the normal and exponential cases, respectively. As we do not observe any hints for a non-linear relationship, the estimation of a linear probability model allows us to determine rules of thumb more easily:

$$y_k = \beta_0 + \beta_{1k} \cdot \lambda + \beta_{2k} \cdot n + \mu_k$$

with

$$y = \begin{cases} 0 & \text{if MOM implies a smaller error} \\ 1 & \text{if ML implies a smaller error} \end{cases}$$

and μ_k as an error term in all k possible combinations of sample size n and λ. Predicted values $\hat{y_k} > 0.5$ imply an advantage of ML- over MOM-estimation. Figure 4.1 illustrates the separating lines between either approach in the half normal and exponential cases. The corresponding parameter estimates are shown in table 4.3.

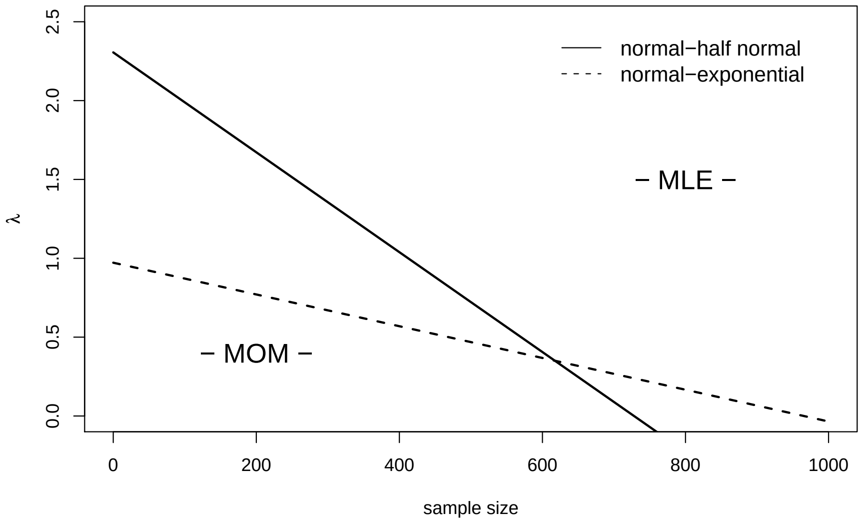

Figure 4.1: Rules of thumb

4.4 An application to German banks

4.4.1 Methodological issues

Inputs and Outputs

In order to illustrate our methodical approach we reviewed the most relevant literature covering bank efficiency estimation on the basis of stochastic frontier analyses. One of the most frequently cited articles in contemporary bank efficiency literature is Mester (1996). The author's procedure reflects the main features of current efficiency analyses via frontier cost functions. Furthermore, resorting to cost functions instead of production functions is inevitable in the case of multiple outputs.

	Estimate	Std. Error	t value	Pr(>|t|)
		normal-half normal		
Intercept	0.0602	0.0900	0.67	0.5062
λ	0.1908	0.0292	6.53	0.0000
n	0.0006	0.0001	4.52	0.0000
		normal-exponential		
Intercept	0.3338	0.0874	3.82	0.0003
λ	0.1710	0.0284	6.02	0.0000
n	0.0002	0.0001	1.33	0.1897

Table 4.3: Parameter estimates rules of thumb

Beside the methodical problems discussed above, we find another key question of empirical bank efficiency estimation in modelling inputs and outputs of the production process[3]. As Girardone et al. (2004) state: *While the multiproduct nature of the banking firm is widely recognized, there is still no agreement as to the explicit definition and measurement of banks' inputs and outputs.* Pasiouras (2008) observes that it is common practice to operationalise bank production according to the fundamental idea of the Intermediation Approach proposed by Sealey and Lindley (1977): The authors proposed a multistage production process of the financial firm, using capital, labour and material to acquire customer deposits in a first step. Lending these funds in a (virtual) second step to deficit spending units, as well as issuing securities and other earning assets involve in general an interest profit. So financial production for intermediation purposes is about adding value to deposits.

Obviously, the use of multiple outputs does not apply to the single-output production functions described above. But, referring to Duality Theory[4], one can, under certain regularity conditions[5], prove the equiva-

[3]Note that we announced in the outline of the research project (chapter 1) that we will refrain from the discussion of 'Adequation' in the course of this and the subsequent methodical chapters, too.

[4]Cp. Beatti and Taylor (1985), chapter 6.

[5]In particular, linear homogeneity and weak concavity in input prices if the implicit production technology is strictly quasi-convex.

lence of indirect cost functions $tc = tc(\mathbf{y}, \mathbf{c})$ and the underlying technology $F(\mathbf{y}, \mathbf{x}) = 0$ with tc total operating costs, y' a vector of outputs, and c' a vector of prices of the inputs x'. The technique to estimate restricted stochastic cost frontiers is virtually the same as the estimation of production frontiers, as the lower stochastic frontier of the 'data cloud' is simply defined by turning the sign of u_i, using the symmetry of v_i:

$$\begin{aligned} \log(tc_i) &= \log\left(g(y_i, c_i); \beta)\right) + v_i + u_i \qquad |v_i \text{ symmetric} \\ \Leftrightarrow -\log(tc_i) &= -\log\left(g(y_i, c_i); \beta)\right) + v_i - u_i \end{aligned}$$

In the context of cost functions efficiency is definded in terms of cost efficiency CE instead of technical efficiency TE. An advantage over considerations of 'pure' technologies is the possibility to evaluate scale economies, i.e. the existence of decreasing average costs in conjunction with the firm's size.

Thus, Mester proposes a basic cost function with outputs being real estate loans, commercial/industrial and other loans as well as loans to individuals. Inputs, input-prices respectively, are prices of labour, physical capital and deposits (borrowed money). Furthermore, she included two bank quality proxies: The average volume of nonperforming loans and the average volume of equity capital. Because of a highly homogeneous dataset no further specific regional/economic distinction had to be drawn. Bos and Kool (2006) investigate small cooperative banks in the Netherlands and confine themselves to just control for a bank-specific solvency measure provided by Rabobank Netherlands, which is added to the cost function. The authors' interpretation of the Intermediation approach imply inputs such as public relations, labour, physical capital and financial capital. Outputs are retail loans, wholesale loans, mortgages and provisions. This is a slight modification of the intermediation idea, but Lang and Welzel (1996) go even further while modelling the outputs short-term and long-term loans to non-banks, loans to banks, bonds, cash and real estate investments, fees and commissions, and revenue from sales of commodities. Obviously, some studies justify more or less distinctive alterations of the value-adding idea by Sealey and Lindley. On the other hand, authors like Altunbas et al. (2000), Perera et al. (2007), Girardone et al. (2004) are able to adopt the classical idea of banking intermediation. Table 4.11 in the appendix shows some different approaches at a glance.

Based on this discussion we put forward the opinion that an adequate idea of banking activity in the course of a simulation study can be deduced by the basic intermediation approach, without adding any control variables.

Our cross sectional data set is obtained from Bankscope[6] and comprises $n = 56$ German commercial banks. The annual statements of account refer to the end of 2005.

In line with the above-mentioned literature we assume banks to minimize total costs rather than operating costs (Pasiouras, 2008) and set up a basic cost function with outputs y_1 interbank loans, y_2 commercial loans, and y_3 securities. Inputs are x_1 fixed assets, x_2 number of employees, and x_3 borrowed funds (deposits). Input prices c_k can be approximated by the ratio of the costs of the inputs x_k to the amount of the particular input. In the case of c_1, c_2 we obtain percentaged values, while c_3 is average cost per employee per year. Table 4.4 shows the descriptive statistics of bank size in terms of total assets ta, inputs $\mathbf{x}$, input prices $\mathbf{c}$ and outputs $\mathbf{y}$.

Var	Description	Mean	St.Dev.	Median
ta	Total assets (BEUR)	3894.079	9043.181	1017.800
tc	Total costs (BEUR)	246.205	615.106	58.150
y_1	Interbank loans (BEUR)	614.568	1374.506	145.650
y_2	Commercial loans (BEUR)	2646.856	7734.673	374.150
y_3	Securities (BEUR)	179.694	561.050	10.200
x_1	Fixed assets (BEUR)	54.552	264.650	5.950
x_2	Employees	645.39	1888.22	142.50
x_3	Borrowed funds (BEUR)	2107.203	5420.554	486.800
c_1	Cost of fixed assets (% depreciation)	0.160	0.089	0.141
c_2	Cost of labour (TEUR/employee)	80.671	64.654	65.621
c_3	Price of funds (% interest expenses)	0.049	0.058	0.031

Table 4.4: Descriptive statistics of inputs, outputs, prices and bank size

[6] Bureau van Dijk, www.bvdep.com.

Shape of the cost function

As for the formal issues, one can state that most authors apply a 'regular' translog cost frontier. In most cases the translog form offers an appropriate balance between flexibility (in price and output elasticities), parameters to estimate, and global fit. The exceptions among the reviewed literature are Altunbas et al. (2000), Altunbas and Chakravarty (2001), Girardone et al. (2004), and Weill (2004), using a Fourier Flexible form with additional trigonometric terms. Otherwise, Fitzpatrick and McQuinn (2005) had to restrict themselves to a simple Cobb-Douglas form due to an insufficient number of observations.

As we, too, work on a small-sized dataset ($n = 56$), we encountered severe multicollinearity in the flexible translog form. So we opt for a log linear Cobb-Douglas cost function (Kumbhakar and Tsionas, 2008). To ensure linear homogeneity in input prices $tc(y, k \cdot c) = k^1 \cdot tc(y, c)$ with $k > 0$, we normalize total costs and input prices by the price of labour c_2 (Lang and Welzel, 1996).

$$\log tc(y,c) = \beta_0 + \sum_{i=1}^{3} \beta_i \log y_i + \sum_{j=1}^{3} \gamma_j \log c_j + v + u \quad \text{s.t.} \quad \sum_{i=j}^{3} \gamma_j = 1$$

The homogeneity-constrained cost frontier results in:

$$\log \frac{tc(y,c)}{c_2} = \beta_0 + \sum_{i=1}^{3} \beta_i \log y_i + \sum_{j=1,3} \gamma_j \log \frac{c_j}{c_2} + v + u$$

4.4.2 Empirical evidence

To our knowledge there is not a single bank efficiency study applying the method of moments estimator we discussed. Conventionally, authors prefer Maximum Likelihood Estimation with the Jondrow et al. (1982) $\exp[-E(u|\varepsilon)]$ estimator mentioned above[7].

As our sample is rather small-sized, we expect the method of moments

[7] Studies applying FRONTIER 4.1 software by Coelli (1996) make implicit use of the alternative point estimator $E(\exp(-u)|\varepsilon)$ (Girardone et al., 2004, Perera et al., 2007, Battese et al., 2000).

approach to deliver more reliable efficiency scores. Table 4.5 shows the results of all scenarios under discussion. Our 'rules of thumb'-indicator gives rather ambivalent recommendations: Predicted values of the binary variable exceeding 0.5 point to MLE application. But especially in the normal-exponential case we are facing values ≈ 0.5. So we should expect rather similar results in both MOM and ML-estimation. In fact, the correlation table 4.6 confirms a $\rho(\widehat{CE}_{MOM}, \widehat{CE}_{MLE}) \approx 0.99$. Mean efficiencies $\frac{1}{n}\sum_i \widehat{CE}_i$ differ only between varying inefficiency distribution assumptions. Moreover, the strong correlations $\rho > 0.95$ between the half normal and exponential approach to inefficiency are in line with the findings of Greene (1990).

	normal-half normal		normal-exponential	
	MLE	MOM	MLE	MOM
$Intercept$	0.760	1.526	2.096	1.819
y_1	0.107	0.157	0.117	0.157
y_2	0.543	0.505	0.552	0.505
y_3	0.055	0.060	0.050	0.060
c_1	0.208	0.233	0.254	0.233
c_3	0.289	0.388	0.418	0.388
λ	2.696	2.804	1.158	0.933
σ_v	0.326	0.325	0.416	0.467
σ_u	0.880	0.913	0.481	0.436
mean $\widehat{CE}$	0.537	0.531	0.655	0.673
rule of thumb	0.608	0.629	0.541	0.503

Table 4.5: Estimation results, all cases

	nhn-mle	nhn-mom	exp-mle	exp-mom
nhn-mle	1.000			
nhn-mom	0.987	1.000		
exp-mle	0.957	0.952	1.000	
exp-mom	0.951	0.952	0.988	1.000

Table 4.6: Correlation table, cost efficiencies

Comparing the parameter estimates, we observe slight differences in the estimated output and price elasticities (note that $\gamma_2 = 1 - \gamma_1 - \gamma_3$). The overall scale economies are calculated as $\epsilon_c = \left(\sum_i \frac{\partial \log tc(y,c)}{\partial \log \beta_i}\right)^{-1}$. As $\epsilon_c > 1$ in all cases, the results hint at increasing returns to scale in the German banking industry.

4.5 Conclusion

We put forward the MOM-approach to stochastic frontiers in bank efficiency analysis. An extensive simulation study confirmed and extended the findings of Coelli (1995) and Olson et al. (1980): Rules of thumb suggest that the MOM-estimation of parametrical frontiers assuming a half normal inefficiency distribution can be favourable in terms of $mse(\widehat{TE}, TE)$ and $mae(\widehat{TE}, TE)$ if the sample size is of medium scale ($\leq$ 700 observations) and inefficiency does not strongly dominate noise ($\lambda < 2$).

In summary, we propose that the method of moment estimation should be considered an alternative to maximum likelihood estimation, especially when efficiency estimation is based on a small sample of banks (Fitzpatrick and McQuinn, 2005). Furthermore, we also recommend the application of MOM-estimation additionally to the ML-procedure even in larger samples as the bias correction based on moments might indicate strong violations of the distributional assumptions (type I and type II errors).

Another practical advantage of MOM-estimation is obvious: Whenever Newton-like numerical optimization is unavailable or fails due to awkward data structure, MOM provides a loophole both robust and easy to implement. A simple two-step procedure (OLS-fitting and bias correction based on estimated residuals) is available within every statistical environment.

4.6 Appendix

4.6.1 Derivatives of the log-likelihood: half-normal

$$\begin{aligned}\ln L(y|\beta,\lambda,\sigma^2) &= n\ln\left(\frac{\sqrt{2}}{\sqrt{\pi}}\right) + n\ln\left(\sigma^{-1}\right)\\ &+ \sum_{i=1}^{n}\ln\left[1-\Phi\left(\left[y_i - x_i'\beta\right]\lambda\sigma^{-1}\right)\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i'\beta)^2\end{aligned}$$

The derivatives are given by

$$\frac{\partial \ln L}{\partial \beta} = -\frac{n}{\sigma^2}\sum_{i=1}^{n}(y_i - x_i'\beta)x_i + \frac{\lambda}{\sigma}\sum_{i=1}^{n}\frac{\phi_i^*}{(1-F_i^*)}x_i$$

$$\begin{aligned}\frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - x_i'\beta)^2\\ &+ \frac{1}{2\sigma^3}\sum_{i=1}^{n}\frac{\phi_i^*}{(1-\Phi_i^*)}(y_i - x_i'\beta)\end{aligned}$$

$$\frac{\partial \ln L}{\partial \lambda} = -\frac{1}{\sigma} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}\frac{\phi_i^*}{(1-\Phi_i^*)}(y_i - x_i'\beta)$$

where

$$\begin{aligned}\phi_i^* &= \phi(\left[\ln y_i - x_i'\beta\right]\lambda\sigma^{-1})\\ \Phi_i^* &= \Phi(\left[\ln y_i - x_i'\beta\right]\lambda\sigma^{-1})\end{aligned}$$

4.6.2 Derivatives of the log-likelihood: exponential

$$\begin{aligned}\ln L(y|\beta, \sigma_u^2, \sigma_v^2) &= -n \ln(\sigma_u) + n\frac{1}{2}\frac{\sigma_v^2}{\sigma_u^2} \\ &+ \sum_{i=1}^{n} \ln \Phi\left(-\frac{1}{\sigma_v}(y_i - x_i'\beta) - \frac{\sigma_v}{\sigma_u}\right) + \sum_{i=1}^{n} \frac{1}{\sigma_u}(y_i - x_i'\beta)\end{aligned}$$

$$\frac{d}{d\beta}\sum_{i=1}^{n} \ln \Phi\left(-\frac{1}{\sigma_v}(y_i - x_i'\beta) - \frac{\sigma_v}{\sigma_u}\right)$$

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{n} \frac{x_i'/\sigma_v \phi\left(-\frac{1}{\sigma_v}(y_i - x_i'\beta) - \frac{\sigma_v}{\sigma_u}\right)}{\Phi\left(-\frac{1}{\sigma_v}(y_i - x_i'\beta) - \frac{\sigma_v}{\sigma_u}\right)} - \sum_{i=1}^{n} x_i'/\sigma_u$$

$$\frac{\partial \ln L}{\partial \sigma_u} = -\frac{n}{\sigma_u} - n\frac{\sigma_v^2}{\sigma_u^3} + \sum_{i=1}^{n} \frac{\sigma_v \phi\left(-\frac{1}{\sigma_v}(y_i - x_i'\beta) - \frac{\sigma_v}{\sigma_u}\right)}{\sigma_u^2 \Phi\left(-\frac{1}{\sigma_v}(y_i - x_i'\beta) - \frac{\sigma_v}{\sigma_u}\right)} - \sum_{i=1}^{n} \frac{1}{\sigma_u^2}(y_i - x_i'\beta)$$

$$\frac{\partial \ln L}{\partial \sigma_v} = n\frac{\sigma_v}{\sigma_u^2} + \sum_{i=1}^{n} \left(\sigma_v^{-2}(y_i - x_i'\beta) - \frac{1}{\sigma_u}\right) \frac{\phi\left(-\frac{1}{\sigma_v}(y_i - x_i'\beta) - \frac{\sigma_v}{\sigma_u}\right)}{\Phi\left(-\frac{1}{\sigma_v}(y_i - x_i'\beta) - \frac{\sigma_v}{\sigma_u}\right)}$$

4.6.3 Tables

n	25	50	75	100	150	250	500	1000
λ	method of moments							
0.5	0.0516	0.0475	0.0464	0.0441	0.0416	0.0383	0.0351	0.0318
1.0	0.0612	0.0565	0.0513	0.0502	0.0464	0.0412	0.0343	0.0284
1.5	0.0588	0.0485	0.0425	0.0390	0.0319	0.0269	0.0226	0.0212
2.0	0.0531	0.0391	0.0319	0.0269	0.0228	0.0192	0.0177	0.0171
2.5	0.0480	0.0299	0.0235	0.0192	0.0166	0.0150	0.0142	0.0137
3.0	0.0424	0.0246	0.0181	0.0158	0.0136	0.0123	0.0115	0.0111
4.0	0.0345	0.0182	0.0131	0.0112	0.0098	0.0087	0.0080	0.0076
5.0	0.0291	0.0150	0.0106	0.0090	0.0077	0.0066	0.0059	0.0055
λ	maximum likelihood							
0.5	0.1370	0.0928	0.0746	0.0639	0.0546	0.0455	0.0359	0.0294
1.0	0.1191	0.0800	0.0667	0.0614	0.0520	0.0414	0.0333	0.0274
1.5	0.0891	0.0578	0.0471	0.0417	0.0323	0.0268	0.0225	0.0211
2.0	0.0654	0.0426	0.0331	0.0269	0.0226	0.0188	0.0176	0.0171
2.5	0.0571	0.0308	0.0229	0.0184	0.0159	0.0145	0.0140	0.0136
3.0	0.0464	0.0240	0.0171	0.0145	0.0127	0.0116	0.0111	0.0110
4.0	0.0348	0.0159	0.0110	0.0095	0.0085	0.0079	0.0076	0.0075
5.0	0.0274	0.0118	0.0082	0.0071	0.0062	0.0059	0.0055	0.0054
λ	advantage							
0.5	MOM	MOM	MOM	MOM	MOM	MOM	MOM	MLE
1.0	MOM	MOM	MOM	MOM	MOM	MOM	MLE	MLE
1.5	MOM	MOM	MOM	MOM	MOM	MLE	MLE	MLE
2.0	MOM	MOM	MOM	MOM	MLE	MLE	MLE	MLE
2.5	MOM	MOM	MLE	MLE	MLE	MLE	MLE	MLE
3.0	MOM	MLE	MLE	MLE	MLE	MLE	MLE	MLE
4.0	MOM	MLE	MLE	MLE	MLE	MLE	MLE	MLE
5.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE

Table 4.7: Mean Square Error normal-halfnormal approach

n	25	50	75	100	150	250	500	1000
λ	method of moments							
0.5	0.0461	0.0441	0.0421	0.0403	0.0387	0.0348	0.0307	0.0274
1.0	0.0566	0.0443	0.0368	0.0339	0.0313	0.0288	0.0278	0.0272
1.5	0.0478	0.0331	0.0287	0.0262	0.0245	0.0232	0.0224	0.0218
2.0	0.0407	0.0274	0.0233	0.0216	0.0201	0.0188	0.0178	0.0172
2.5	0.0372	0.0229	0.0198	0.0184	0.0167	0.0154	0.0143	0.0136
3.0	0.0338	0.0207	0.0176	0.0160	0.0146	0.0129	0.0119	0.0111
4.0	0.0308	0.0182	0.0150	0.0132	0.0115	0.0099	0.0087	0.0078
5.0	0.0305	0.0168	0.0138	0.0116	0.0098	0.0082	0.0068	0.0059
λ	maximum likelihood							
0.5	0.0612	0.0515	0.0469	0.0446	0.0413	0.0363	0.0311	0.0275
1.0	0.0692	0.0490	0.0393	0.0353	0.0314	0.0288	0.0275	0.0272
1.5	0.0535	0.0339	0.0279	0.0249	0.0232	0.0222	0.0217	0.0215
2.0	0.0434	0.0250	0.0204	0.0189	0.0177	0.0171	0.0167	0.0165
2.5	0.0358	0.0193	0.0159	0.0148	0.0139	0.0134	0.0130	0.0128
3.0	0.0297	0.0156	0.0130	0.0120	0.0113	0.0107	0.0103	0.0102
4.0	0.0239	0.0111	0.0091	0.0084	0.0076	0.0072	0.0069	0.0068
5.0	0.0206	0.0088	0.0069	0.0062	0.0055	0.0052	0.0049	0.0048
λ	advantage							
0.5	MOM	MOM	MOM	MOM	MOM	MOM	MOM	MOM
1.0	MOM	MOM	MOM	MOM	MOM	MOM	MLE	MOM
1.5	MOM	MOM	MLE	MLE	MLE	MLE	MLE	MLE
2.0	MOM	MLE	MLE	MLE	MLE	MLE	MLE	MLE
2.5	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
3.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
4.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
5.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE

Table 4.8: Mean Square Error normal-exponential approach

n	25	50	75	100	150	250	500	1000
λ	method of moments							
0.5	0.2035	0.2022	0.1992	0.1963	0.1933	0.1874	0.1850	0.1801
1.0	0.2100	0.2013	0.1931	0.1923	0.1938	0.1934	0.1965	0.1969
1.5	0.1896	0.1817	0.1841	0.1845	0.1895	0.1931	0.1967	0.1982
2.0	0.1748	0.1745	0.1775	0.1787	0.1858	0.1910	0.1953	0.1966
2.5	0.1680	0.1685	0.1734	0.1775	0.1830	0.1871	0.1935	0.1945
3.0	0.1639	0.1670	0.1724	0.1746	0.1821	0.1875	0.1945	0.1975
4.0	0.1554	0.1622	0.1665	0.1744	0.1829	0.1894	0.1969	0.1995
5.0	0.1532	0.1592	0.1660	0.1703	0.1816	0.1906	0.1980	0.2011
λ	maximum likelihood							
0.5	0.2537	0.2280	0.2139	0.2078	0.1934	0.1846	0.1785	0.1756
1.0	0.2291	0.1960	0.1840	0.1800	0.1763	0.1736	0.1738	0.1736
1.5	0.1834	0.1614	0.1555	0.1536	0.1513	0.1508	0.1503	0.1502
2.0	0.1533	0.1381	0.1343	0.1317	0.1309	0.1303	0.1297	0.1295
2.5	0.1321	0.1197	0.1170	0.1165	0.1142	0.1136	0.1128	0.1124
3.0	0.1176	0.1058	0.1049	0.1032	0.1018	0.1005	0.0997	0.0995
4.0	0.0964	0.0861	0.0841	0.0841	0.0825	0.0815	0.0808	0.0804
5.0	0.0858	0.0724	0.0713	0.0707	0.0700	0.0689	0.0679	0.0674
λ	advantage							
0.5	MOM	MOM	MOM	MOM	MOM	MLE	MLE	MLE
1.0	MOM	MLE	MLE	MLE	MLE	MLE	MLE	MLE
1.5	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
2.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
2.5	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
3.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
4.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
5.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE

Table 4.9: Mean Average Error misspecification 1

n	25	50	75	100	150	250	500	1000
λ	method of moments							
0.5	0.1451	0.1409	0.1382	0.1377	0.1351	0.1331	0.1300	0.1281
1.0	0.1884	0.1813	0.1742	0.1728	0.1690	0.1638	0.1559	0.1479
1.5	0.1964	0.1823	0.1739	0.1698	0.1608	0.1538	0.1474	0.1450
2.0	0.1962	0.1758	0.1664	0.1587	0.1525	0.1466	0.1433	0.1412
2.5	0.1935	0.1677	0.1590	0.1510	0.1454	0.1423	0.1392	0.1379
3.0	0.1902	0.1628	0.1514	0.1472	0.1434	0.1391	0.1361	0.1352
4.0	0.1829	0.1571	0.1475	0.1430	0.1389	0.1355	0.1327	0.1320
5.0	0.1773	0.1537	0.1445	0.1419	0.1367	0.1342	0.1316	0.1304
λ	maximum likelihood							
0.5	0.1688	0.1548	0.1488	0.1469	0.1425	0.1392	0.1351	0.1322
1.0	0.2117	0.1959	0.1865	0.1837	0.1780	0.1706	0.1612	0.1510
1.5	0.2124	0.1921	0.1798	0.1749	0.1644	0.1554	0.1472	0.1440
2.0	0.2044	0.1781	0.1666	0.1554	0.1485	0.1410	0.1363	0.1338
2.5	0.1954	0.1612	0.1511	0.1408	0.1322	0.1289	0.1252	0.1235
3.0	0.1838	0.1489	0.1348	0.1284	0.1234	0.1180	0.1152	0.1140
4.0	0.1669	0.1312	0.1168	0.1107	0.1056	0.1022	0.0994	0.0982
5.0	0.1525	0.1168	0.1036	0.0982	0.0932	0.0903	0.0877	0.0862
λ	advantage							
0.5	MOM	MOM	MOM	MOM	MOM	MOM	MOM	MOM
1.0	MOM	MOM	MOM	MOM	MOM	MOM	MOM	MOM
1.5	MOM	MOM	MOM	MOM	MOM	MOM	MLE	MLE
2.0	MOM	MOM	MOM	MLE	MLE	MLE	MLE	MLE
2.5	MOM	MLE	MLE	MLE	MLE	MLE	MLE	MLE
3.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
4.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
5.0	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE

Table 4.10: Mean Average Error misspecification 2

Inputs	Outputs	Reference
Altunbas and Chakravarty (2001)		
labour, total funds, physical capital	all types of loans, total aggregate securities, off-balance sheet activities	Intermediation (varied)
Battese et al. (2000)		
public loans, guarantees, deposits, number of branches, value of inventories	costs of labour use	Input-requirement model
Bos and Kool (2006)		
public relations, labour, housing, physical capital	retail loans, wholesale loans, mortgages, provisions	Intermediation (varied)
Ferrier and Lovell (1990)		
employees, occupancy costs, materials	demand deposit accounts, time deposit accounts, real estate loans, real estate loans, installment loans, commercial loans	Production
Fitzpatrick and McQuinn (2005)		
labour, physical capital, financial capital	consumer/commercial/other loans, non-interest revenue	Intermediation (varied)
Girardone et al. (2004)		
employees, total customer deposits, total fixed assets	total customer loans, other earning assets	Intermediation
Lang and Welzel (1996)		
employees, fixed assets, deposits	short-term and long-term loans to non-banks, loans to banks, bonds/cash/real estate investments, fees and commissions, revenue from sales of commodities	Intermediation (varied)
Mester (1996)		
labour, physical capital, deposits	real estate loans, commercial/industrial/government/... loans, loans to individuals	Intermediation
Pasiouras (2008)		
total deposits, total costs, equity	loans, other earning assets, non-interest income	Intermediation (varied)
Perera et al. (2007)		
funds, labour, capital	net total loans, other earning assets	Intermediation
Weill (2004)		
labour, physical capital, borrowed funds	loans, investment assets	Intermediation

Table 4.11: Input and output-modelling in selected bank efficiency studies

Chapter 5

5 Cost efficiency trends in European and US Banking

5.1 Introduction

In recent decades we have observed a vast number of articles covering bank efficiency assessments. Especially studies dealing with international bank data mostly end in an efficiency ranking on the basis of mean efficiency scores in the respective country. We tend to call this a 'static' ranking, since the conventional methodological treatment of the time dimension of the underlying panel structure implies the assumption of a common trend in efficiency for all banks in the sample. So the relative position of the countries' banking systems remains unaffected over the years. To our knowledge, the opportunity of 'dynamic' rankings of banking systems per country *and* year has never been performed in literature, although the methods are available, e.g. Cuesta (2000), Cornwell et al. (1990). So we propose that the abandonment of the assumption of a common trend in efficiency in favour of country-specific trends may reveal new findings.

Since the first cross sectional approaches to a Stochastic Frontier Analysis (SFA) by Aigner et al. (1977) and Meeusen and van de Broeck (1977) the methodology is now technically mature. We picked out two entirely different methods in panel SFA and developed the estimation of group-specific efficiency trends. With respect to the SFA methods implemented within most statistical software packages there exists only the option of estimating a common trend parameter for the whole sample. This leads to an unchanged efficiency structure over time for all firms in the data set: Consequently, the efficiency ranking is static. On the other hand, partitioning the data in subsamples per group allows for group-specific trends, but one loses the possibility of group rankings against the background of the same benchmark. The methods we put forward combine the option of estimating group-specific time trend parameters – allowing for more heterogeneity – and the validity of inter-group efficiency comparisons.

For these purposes our study is composed as follows: In chapter 5.2, we provide an overview of a sample of recent and most cited studies on international bank efficiency comparisons. Eventually, we try to stick as closely as possible to the established procedures in bank efficiency studies to ensure the basic comparability of our formal results with preceding studies. In chapter 5.3, we describe two SFA approaches (fixed and random effects) and develop the possibility to estimate group-specific time trends within the panel structure. The aim is to motivate the implementation of the (numerical) optimization procedures within statistical programming environments. Chapter 5.4 contains descriptive statistics on the compilation of our database, comprising Bankscope and OECD data. In chapter 5.5, we discuss the dynamic efficiency ranking we estimated. Chapter 5.6 finally contains the conclusion.

5.2 Literature overview

Despite the fact that the majority of efficiency studies focus on individual countries, several comparative studies of efficiency in financial institutions have been undertaken. Among them we reviewed Al-Sharkas et al. (2008), Altunbas et al. (2001), Bos and Schmiedel (2003), Casu and Molyneux (2003), Casu and Girardone (2006), Dietsch and Lozano-Vivas (2000), Girardone et al. (2009), Goddard et al. (2004), Lozano-Vivas et al. (2002), Oliveira and Tabak (2005), Pastor et al. (1997), Pastor and Serrano (2006), Weill (2004). Table 5.8 in the appendix shows the basic information on countries and years covered as well as methods applied.

Obviously, European and US banking systems on the turn of the millenium constitute the focus of interest. Especially the ongoing process of European financial integration has aroused the need for an impartial 'yardstick' regarding the potential gains on banking competition and costs.

With respect to methodological issues, we counted an approximately balanced number of applications of either parametrical SFA or non-parametrical Data Envelopment Analysis (DEA), and only few applications of the Distribution Free Approach (DFA) (Berger, 1993). As the underlying motivation for bank efficiency studies is basically twofold – besides economic questions several authors discuss the implications of different methodological approaches – we sometimes encounter simultaneous application of multiple methods (Al-Sharkas et al., 2008, Lozano-Vivas et al., 2002, Weill, 2004).

The key to bank efficiency estimation is the definition of typical banking activities, i.e. the delienation for some kind of production process. The standard approach is the assumption of financial transformation processes described by the 'idea of intermediation' (Sealey and Lindley, 1977)[1]. It concerns a multistage production process of the financial firm, using capital, labour and material to acquire customer deposits in a first step. Lending these funds in a (virtual) second step to deficit spending units and issuing securities and other earning assets involve in general an interest profit. So financial production for intermediation purposes is about adding value to deposits. As most bank efficiency studies are based on Bankscope balance sheet data, the authors have access to the same set of multiple input and output variables. In particular, we most frequently find as outputs (1) all kinds of loans and (2) securities/investments, and sometimes (3) off-balance sheet items (OBS), e.g. contingent liabilities. Although OBS activities do not accord with the classical understanding of financial intermediation processes, most authors see the growing importance of considering non-traditional banking activities[2]. Inputs are (1) labour (number of employees or personnel expenses),(2) physical capital, and (3) borrowed funds/deposits (Al-Sharkas et al., 2008, Altunbas et al., 2001, Bos and Schmiedel, 2003, Casu and Molyneux, 2003, Casu and Girardone, 2006, Girardone et al., 2009, Pastor and Serrano, 2006, Weill, 2004).

Alternatively, some studies refer to the 'added value' approach by

[1]This reasoning is sometimes referred to as the asset approach to defining bank output since funds intermediation is the focus rather than deposit service production (Hancock, 1991, p.16).

[2]Casu and Girardone (2005) showed that omitting OBS activities may lead to biased productivity estimates.

Berger et al. (1993). This implies treating a balance sheet item (whether liability or asset) as an output, if the returns on the product are higher than its opportunity cost (Pastor et al., 1997). As such detailed information on banks' activities and opportunity costs are typically unavailable, most authors resort to the pioneering results by Berger and Humphrey (1992): Thus, (1) all kinds of deposits (demand, savings and time), (2) loans, and (3) other productive assets, i.e. investments or deposits with banks, generate more added value and ought to be treated as outputs. Inputs are (1) personnel expenses and (2) other non-interest expenses (Pastor et al., 1997, Lozano-Vivas et al., 2002). Dietsch and Lozano-Vivas (2000) suggest inputs (1) labour, (2) physical capital $\equiv$ (capital equipment + occupancy expenses)/fixed assets and (3) a financial factor $\equiv$ total interest expenses/total interest bearing liabilities.

Application of SFA methods allows for the presence of just one dependent variable. In the context of multiple-output multiple-input production processes, we can deal with the problem by resorting to cost or profit functions. Incorporating the conditions of Duality Theory[3], cost functions contain the full information about the underlying production technology. Authors applying DEA can circumvent these issues by directly setting up weighted input-output-ratios (Casu and Molyneux, 2003, Casu and Girardone, 2006, Lozano-Vivas et al., 2002, Pastor et al., 1997, Pastor and Serrano, 2006).

Apart from these two standard procedures (intermediation approach and added value approach), we observe two exceptions within the reviewed literature: Goddard et al. (2004) and Oliveira and Tabak (2005) do not necessarily link bank efficiency with considerations about bank production processes. They alternatively take common profitability measures, i.e. the return on equity (ROE), as a reference. Goddard et al. (2004) regress the ROE on total assets, the share of off-balance sheet activities, the capital-to-asset ratio as well as dummies identifying public or commercial banks. As the authors aim at reporting a significant relationship between dependent and explanatory variables, the application of OLS and common panel methods serves the purpose. Oliveira and Tabak (2005) accessed Datastream and Bloomberg financial data to constitute a DEA frontier

[3]Cp. Beatti and Taylor (1985), chapter 6.

assessing banks' ROE given country-specific stock market characteristics. As their study covers an impressive number of 41 countries worldwide the authors can classify their results according to the specific market development in several stages.

Basically, it is difficult to compare results of international bank efficiency studies. As the obtained DEA and SFA efficiency scores are only weakly correlated between each other, even though based on identical data (Weill, 2004), we have to assume the absence of any coherence among studies applying different methods on different data. Another reason for noticeable differences is the treatment of country-specific technologies. On the one hand, estimation of separate frontiers for each country perfectly accounts for technological and environmental particularities (Weill, 2004). But unfortunately this prevents the cross-national comparison of efficiency scores with the objective of country rankings. On the other hand, estimation of a common frontier is the traditional approach. Assuming an identical benchmark technology across different countries is obviously a violation of sample homogeneity as it tends to generate too much inefficiency: First, Casu and Molyneux (2003) confirm that most of the efficiency differences found across European banking systems are due to country-specific aspects of the banking technology. And second, Bos and Schmiedel (2003) substantiate the fact that EU banking institutions do not always have access to the same benchmark technology. Thus, Dietsch and Lozano-Vivas (2000) are the first to suggest selecting a set of environmental variables to account for country-specific conditions[4]. Unfortunately, Lozano-Vivas et al. (2002) confirm that no theoretical studies exist as for the choice of the appropriate environmental control variables. In consequence, most authors rely on previous empirical studies or data availability. A typical set of commonly used environmental variables can be salary per capita, population density, density of demand, the Herfindahl-Index indicating the degree of competition among banks, intermediation ratio, branch density, ownership characteristics or even the level of financial development according to the World Bank Financial

[4] Berger (2007) confirms that Dietsch and Lozano-Vivas (2000) is the first 'improved' study on the turn of the century specifying better controls for differences in economic environments.

Structure Database[5]. Alternatively, a simple country-specific dummy variable accounts for most of the observed heterogeneity, but eliminates any differences in inefficiency levels.

Besides country-specific heterogeneity, we additionally encounter differences due to specialization and/or ownership characteristics. On this account, most studies deal with homogenized bank data sets: Either adjusted by similar bank size in terms of total assets (Pastor and Serrano, 2006), or considering only commercial banks assuming a common profit-maximizing behaviour (Pastor et al., 1997, Lozano-Vivas et al., 2002, Bos and Schmiedel, 2003, Goddard et al., 2004), or incorporating a specialization dummy to distinguish commercial/private banks from savings/public banks (Casu and Molyneux, 2003). As Oliveira and Tabak (2005) use stock and market risk data, the authors confine themselves to market disciplined institutions. Pastor and Serrano (2006) take an interesting approach, regardless of the banks' legal form: By means of a cluster analysis, the institutions 'automatically' fall into four groups (1) mortgage and intermediation banking, (2) retail banking, (3) wholesale banking and (4) universal banking. The authors discuss in detail the implications of adherence to a particular group.

Given the vast number of studies on bank efficiency it is rather impossible to make general statements about the methods that have been applied hitherto. We reviewed many more studies than our selection presented here comprises. Insofar, we have good reason to assume that most studies on international efficiency comparisons end in a static efficiency ranking. Hence, we intend to put forward in the following section two innovative methods that allow *dynamic* efficiency rankings per year and country. As for the operationalization of financial production processes we try to stick to the well-established standard procedure, in close dependence on the most recent study by Girardone et al. (2009).

[5] Girardone et al. (2009): The financial structure index quantifies the degree of stock market orientation of a financial system according to the three criteria relative stock market capitalization, value of stocks traded in relation to private loans and value of stocks traded multiplied with relative overhead costs in the respective countries.

5.3 SFA methodology

5.3.1 The basic idea of productive efficiency

The derivation of cross-sectional methods in SFA has often been performed in literature. The standard references are Aigner et al. (1977) and Meeusen and van de Broeck (1977). So we just give a short introduction to the idea of the model. For a comprehensive discussion of a maximum likelihood approach versus a method of moments approach in the context of a half-normally or exponentially distributed inefficiency term we refer to chapter 4.

Let us assume cross sectional data for n units indexed by i $(i = 1, ..., N)$ using K $(k = 1, ..., K)$ different inputs contained in the input vector x_i to produce a single output y_i. The fundamental idea of stochastic frontier technical efficiency can be formalized as the ratio of realized output, given a specific set of inputs, to maximum attainable output (Coelli et al., 2005, p. 244):

$$TE_i = \frac{y_i}{y_i^*} = \frac{\exp(g(x_i;\beta)) \cdot \exp(v_i) \cdot \exp(-u_i)}{\exp(g(x_i;\beta)) \cdot \exp(v_i)} = \exp(-u_i) \in (0, 1]$$

with $g(x_i;\beta)$ as a log-linear production function (e.g. Cobb-Douglas, Translog or Fourier Flexible Form), v_i as random noise $v_i \sim N(0;\sigma_v^2)$ and $u_i \geq 0$ representing inefficiency. y_i^* is the maximum attainable output for unit i given x_i.

The estimation of the deterministic part of the frontier function $g(x_i;\beta)$ cannot be performed by OLS, as the error term is two-part and asymmetrically distributed: $\varepsilon_i \equiv v_i - u_i$ with $E(\varepsilon) \leq 0$. In practice, the maximum likelihood estimation leads to slope parameters similar to OLS and a vertically shifted intercept.

The fundamental issue to any SFA estimation is now the disentanglement of a two-component error term $\varepsilon_i = v_i - u_i$. As we only observe a single deviation $\hat{\varepsilon}_i$ per firm i, it is virtually impossible to generate two estimators $\hat{v}_i$ and $\hat{u}_i$ at the same time. An increasing sample size $N \to \infty$ does not improve the accuracy of every $\hat{u}_i$.

For this reason, the application of cross-sectional SFA has to rely on conditional expected values of inefficiency $E(u|\varepsilon)$. Only multiple observations on the same firm provide the opportunity to overcome this information problem. This is the case in a panel data environment with T_i observations on every firm-specific inefficiency score u_i.

We will not confine ourselves to the most simple case of time-invariant inefficiency. In its most general form, the production model in terms of panel data can be written as

$$y_{it} = \beta_{0it} + \sum_{k=1}^{K} \beta_{kit} x_{kit} + v_{it} - u_{it} \tag{5.1}$$

with $i = 1, ..., N$ firms, $t = 1, ..., T_i$ years per i, y_{it} firm i's output in t, x_{kit} firm i's inputs $k = 1, ..., K$ in t, v_{it} normal noise and $u_{it} \geq 0$ firm-specific inefficiency in t.

Obviously, the number of observations $\sum_{i=1}^{N} T_i$ is insufficient to estimate $\sum_{i=1}^{N} T_i$ intercepts β_0, $\left(\sum_{i=1}^{N} T_i\right) \cdot K$ slope parameters as well as additional parameters characterizing the distributions of v and u. In consequence, certain restrictions are to be imposed on the structure of the model.

5.3.2 Fixed effects with time trend

A model with time-variant efficiency and common slope parameters for all t and i is given as (Kumbhakar and Lovell, 2003, p. 108):

$$y_{it} = \beta_{0t} + \sum_{k=1}^{K} \beta_k x_{kit} + v_{it} - u_{it} \tag{5.2}$$

with β_{0t} as intercept of the production frontier for all i in t. A simple approach to the estimation of this model is the 'fixed effects interpretation':

$$\beta_{0it} \equiv \beta_{0t} - u_{it} \tag{5.3}$$

We simply defined the firm-specific amount of inefficiency in t as vertical distance between a true but unobserved β_{0t} constituting the efficient upper (production) boundary and the firm specific effect in t. This leads to

$$y_{it} = \beta_{0it} + \sum_{k=1}^{K} \beta_k x_{kit} + v_{it} \tag{5.4}$$

So u_{it} is no more part of the error term, but included in a firm- and time-specific intercept ('effect'). For that reason, no particular random distribution for u_{it} has to be assumed. Thus we can speak of a semiparametric treatment of inefficiency[6].

In order to further limit the number of parameters, most authors impose a certain deterministic structure on the varying intercept term over time. For example, Cornwell et al. (1990, p. 192) propose a unit-specific time trend of the form

$$\beta_{0it} = \beta_{0i} + \omega_{1i} t + \omega_{2i} t^2 \tag{5.5}$$

Estimation of the model by either the within transformation or dummy matrices is straightforward. But, as we deal with panel data where the ratio N/T is rather large[7], we simplify this idea to the estimation of country-specific ω_{1l}, ω_{2l}, with every i belonging to one country $l = 1, ..., L$, and with $L \ll N$:

$$\beta_{0it} = \beta_{0i} + \omega_{1l} t + \omega_{2l} t^2 \quad \text{for } i \text{ in } l \tag{5.6}$$

So the final model is

$$y_{it} = \beta_{0i} + \omega_{1l} t + \omega_{2l} t^2 + \sum_{k=1}^{K} \beta_k x_{kit} + v_{it} \quad \text{for } i \text{ in } l \tag{5.7}$$

In practice, the OLS estimation can be performed by construction of appropriate time-index matrices of dimension $\sum_{i=1}^{N} T_i \times L$. We obtain L time trend functions and N intercepts $\hat{\beta}_{0i}$. The time-dependent $\hat{\beta}_{0it}$ can be recovered according to equation (5.6). Now firm-specific inefficiencies $\hat{u}_{it}$ can be estimated (equation 5.3). We set the most efficient firm per year as 100% efficient. This is justifiable in the marginal case $N \to \infty$.

[6] An example of a non-parametric treatment is the Data Envelopment Analysis (DEA). In DEA models, a random term does not exist at all.

[7] In cases the dummy matrices exceed computer capacity.

Based on this assumption, Schmidt and Sickles (1984) proposed the relative inefficiency estimator[8]:

$$\begin{aligned}
\hat{\beta}_{0t} &= \max_i(\hat{\beta}_{0it}) && \text{for } t = 1, ..., \max T_i \\
\hat{u}_{it} &= \hat{\beta}_{0t} - \hat{\beta}_{0it} && \text{for } i = 1, ..., N \text{ and } t = 1, ..., \max T_i \\
\widehat{TE}_{it} &= \exp(-\hat{u}_{it}) && \text{for multiplicative production frontiers}
\end{aligned}$$

This is the same as

$$\widehat{TE}_{it} = \frac{\exp(\hat{\beta}_{0it})}{\exp(\hat{\beta}_{0t})} \in (0, 1] \tag{5.8}$$

In practice (using data sets smaller than infinity), the `max` operator comes along with a strong sensitivity towards outliers. This results in very poor efficiency scores for the remaining units. For this reason, we recommend replacing the `max` operation with a high quantile, for example 99% in the case of production frontiers. The efficiency scores for units excessing frontier production have to be adjusted to a 100% score.

So we can conclude: Panel stochastic frontier analysis based on the fixed effects approach leads to consistent estimators for efficiency scores $\widehat{TE}_{it}$ as $T_i \to \infty$ and $N \to \infty$. No random distribution for inefficiency u_{it} has to be defined and it allows for correlation between effects and regressors. If desired, inference to u_{it} in so-called MCB-models (*multiple comparisons with the best*) can be derived by bootstrapping procedures (Greene, 2008, p.186).

5.3.3 Random effects with time trend

Kumbhakar (1990) developed a flexible approach to technical inefficiency on the basis of a Random Effects approach[9] which can handle different types of time behavior, as well as time-*in*variant efficiency as a special case.

[8] The authors originally discussed the case of time-invariant inefficiency

[9] I.e., a random distribution for the inefficiency term is explicitely defined.

In equation 5.2, u_{it} takes the form

$$u_{it} = \gamma(t) \cdot \tau_i \tag{5.9}$$

with $\tau_i \sim iidN^-(0, \sigma_\tau^2)$ a half normal firm-specific factor. $\gamma(t)$ is a deterministic function of time and is considered to take on a form with two parameters

$$\gamma(t) = \frac{1}{1 + \exp(\omega_1 t + \omega_2 t^2)} \tag{5.10}$$

Obviously, $\gamma(t)$ is bounded by the positive interval $(0, 1)$. Multiplication by $\tau_i \leq 0$ leads to inefficiency levels $u_{it} \leq 0$. In the case of $\omega_1 = \omega_2 = 0$, inefficiency does not vary over time. As the rate of productivity change is the same for all producers, the efficiency ranking between all i remains unaltered for $t = 1, ..., T_i$. Thus, this approach turns out to be the precursor of the more familiar Battese and Coelli (1992) model. The authors assume a similar deterministic trend function:

$$\begin{aligned} u_{it} &= \eta(t) \cdot \tau_i \\ &= \Big(1 + \eta_1(t - T) + \eta_2(t - T)^2\Big) \cdot \tau_i \end{aligned} \tag{5.11}$$

with T the last year in the sample and τ_i allowed to be truncated normal with $E(\tau_i) \neq 0$. The popularity of this alternative model – we found the most recent application in Girardone et al. (2009) – arises mainly from the fact that the authors published a freely available computer program for estimation[10]. Unfortunately, the program is limited to the estimation of just one time trend parameter η_1. For this reason we continue our description of the Kumbhakar (1990) approach.

The production function can be written as

$$y_{it} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{kit} + v_{it} + \gamma(t)\tau_i \tag{5.12}$$

OLS estimation is inappropriate, because the expected value of the composed error term $v_{it} + \gamma(t)\tau_i$ is nonzero. The expected bias $E(v_{it} +$

[10]http://www.uq.edu.au/economics/cepa/frontier.htm

$\gamma(t)\tau_i) = \gamma(t) \cdot E(\tau_i)$ may be corrected by a Method of Moments approach (Kumbhakar and Lovell, 2003, p. 113), but Kumbhakar (1990) suggests Maximum Likelihood estimation.

Let $\varepsilon_{it} = v_{it} + u_{it}$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, ..., \varepsilon_{iT_i})'$ denote the resulting vector of firm-specific residuals and T_i the number of years firm i is observed (for an unbalanced panel). Since the v_{it} are *iid* and independent of τ_i, the joint density of $(v_{i1}, ..., v_{iT_i}, \tau_i)'$ is

$$\begin{aligned} f(v_{i1}, ..., v_{iT_i}, \tau_i) &= f(\boldsymbol{v_i}) \cdot f(\tau_i) = f(\boldsymbol{\varepsilon_i} - \boldsymbol{u_i}) \cdot f(\tau_i) \\ &= f(\boldsymbol{\varepsilon_i} - \gamma(t)\tau_i) \cdot f(\tau_i) \end{aligned}$$

with $\boldsymbol{u}_i = (u_{i1}, ..., u_{iT_i})'$. This is also the joint density of $f(\boldsymbol{\varepsilon}_i, \tau_i)$. The marginal density $f(\boldsymbol{\varepsilon}_i)$ can be obtained by integration with respect to the range for τ_i, which is (Kumbhakar, 1990, p. 205)

$$\begin{aligned} f(\boldsymbol{\varepsilon}_i) &= \int_{-\infty}^{0} f(\boldsymbol{\varepsilon}_i, \tau_i) d\tau_i \\ &= \frac{2\sigma_i^* \exp(-a_i^*/2)}{(2\pi)^{T_i/2} \sigma_v^{T_i} \sigma_\tau} \cdot \Phi(-\mu_i^*/\sigma_i^*) \end{aligned}$$

where

$$\begin{aligned} \sigma_i^* &= \frac{\sigma_v \sigma_\tau}{\left(\sigma_v^2 + \sigma_\tau^2 \sum_t^{T_i} \gamma^2(t)\right)^{1/2}} \\ \mu_i^* &= \frac{\sum_t^{T_i} \gamma(t)\varepsilon_{it}}{\sigma_v^2} \cdot \sigma_i^{*2} \\ a_i^* &= \frac{1}{\sigma_v^2} \left[\sum_t^{T_i} \varepsilon_{it}^2 - \sigma_\tau^2 \left(\sum_t^{T_i} \gamma(t)\varepsilon_{it} \right)^2 \left(\sigma_v^2 + \sigma_\tau^2 \sum_t^{T_i} \gamma^2(t) \right)^{-1} \right] \end{aligned}$$

The objective function is the log-likelihood function for all producers

$i = 1, ..., N$, each observed in $t = 1, ..., T_i$.

$$\ln L = \sum_{i=1}^{N} \ln(2\sigma_i^*) - \frac{1}{2}\sum_{i=1}^{N} a_i^* - \left(\frac{1}{2}\ln(2\pi) + \ln\sigma_v\right) \cdot \sum_{i=1}^{N} T_i$$
$$- N\ln\sigma_\tau + \sum_{i=1}^{N} \ln\left[1 - \Phi\left(\frac{\mu_i^*}{\sigma_i^*}\right)\right] \tag{5.13}$$

Replacing[11] $\varepsilon_{it} = y_{it} - \beta_0 - \sum_k \beta_{k=1}^K x_{kit}$ and optimization of equation (5.13) by standard procedures yield estimated values $\hat{\beta}_0, \hat{\boldsymbol{\beta}}_k, \hat{\omega_1}, \hat{\omega_2}, \hat{\sigma}_v$ and $\hat{\sigma}_\tau$. Firm-specific efficiency scores can be recovered by deriving the expectation of $\tau_i|\boldsymbol{\varepsilon}_i$ with respect to the conditional density:

$$f(\tau_i|\boldsymbol{\varepsilon}_i) = \frac{f(\boldsymbol{\varepsilon}_i, \tau_i)}{f(\boldsymbol{\varepsilon}_i)} = \frac{1}{\sqrt{2\pi}\sigma_i^*} \cdot \frac{\exp\left(-\frac{1}{2\sigma_i^{*2}}(\tau_i - \mu_i^*)^2\right)}{\Phi(-\mu_i^*/\sigma_i^*)} \tag{5.14}$$

The extended Jondrow et al. (1982)-like estimator is

$$\hat{\tau}_i = E(\tau_i|\boldsymbol{\varepsilon}_i) = \mu_i^* - \sigma_i^* \cdot \frac{\phi(\mu_i^*/\sigma_i^*)}{\Phi(-\mu_i^*/\sigma_i^*)} \tag{5.15}$$

The time-variant inefficiency scores are

$$\hat{\boldsymbol{u}}_i = \hat{\boldsymbol{\gamma}} \cdot \hat{\tau}_i \tag{5.16}$$

with $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}(1), ..., \hat{\gamma}(\max T_i))$. This leads to technical efficiencies in a production frontier environment

$$\widehat{TE}_{it} = \exp[E(\tau_i|\boldsymbol{\varepsilon}_i) \cdot \hat{\gamma}(t)] \tag{5.17}$$

ML-estimation of the parameters provides consistent estimators of $\hat{\boldsymbol{u}}_i$ as $T_i \to \infty$.

Up to this point, we have dealt with the case of a single common efficiency trend for the whole sample. Just as we modified the classical fixed effects approach to capture group-specific trend parameters, we

[11] The Jacobian of the transformation from (v, u) to (ε, u) has $det = 1$. Cp. Greene (2008, p. 116).

intend to develop the Kumbhakar (1990) model to handle group-specific trends, too[12].

$$\gamma_l(t) = \frac{1}{1 + \exp(\omega_{1l} t + \omega_{2l} t^2)} \quad \text{for } i \text{ in } l \tag{5.18}$$

The maximum likelihood procedure described above does not change at all. By means of the same time index matrices applied in subsection 5.3.2, one can ensure that every firm i is assigned to the respective group l.

5.4 Descriptive statistics of the database

We will now turn to our exemplifying application: the estimation and assessment of bank efficiency scores. As indicated in chapter 5.2, we deal with international data on banks' balance sheets and income statements as reported in the familiar Bankscope database[13]. The working data set consists of commercial, savings and cooperative banking institutions in eight OECD countries holding a major occurrence in the Bankscope database. In particular, it comprises Switzerland (CH), Germany (DE), Spain (ES), France (FR), Italy (IT), Sweden (SE), Great Britain (GB) and the USA (US). The length of the panel is eight years, from 2000 to 2007. So we cover the most recent data in all studies reviewed.

5.4.1 Data preparation and imputation

The major advantage of the Bankscope database is the aggregation of nation-specific balance sheet items according to a global scheme, as well as the disclosure in a common currency, e.g. US-Dollar (USD). Inevitably, this functionality goes along with a diminished possibility to inspect detailed information. Moreover, we encountered fragmentarily administered data. To avoid losing too many observations we successfully tried to reconstruct some of the missing data. For this reason we

[12] This idea is based on the approach by Cuesta (2000), extending the Battese and Coelli (1992) model by unit-specific trend parameters η_i.

[13] Bureau van Dijk.

reconstructed basic balance sheet and income statement schemata on the basis of the available variables. In cases where just one item in assets or liabilities is missing, we were able to unambiguously recover more than 3.000 missing observations using unambiguous balance sheet identities. Thereafter, we prepared a data set of all interesting variables. In cases where observations on just one single variable per unit were missing, a $k-$nearest-neighbour imputation algorithm was finally applied[14]. It compares the units on the basis of the existing observations, identifying the $k = 10$ units that most resemble each other. Eventually, the algorithm calculates the missing value as the inter- or extrapolation of the k existing observations. We performed this procedure in country-specific subsets.

After deleting implausible values (we treat observations ≤ 0 in major balance sheet items, e.g. total deposits, total costs, equity etc. as implicitly missing), and excluding banks observed in only one year, we find the number of banks per year and country given in table 5.1. In figure 5.4 in the appendix we give an overview of how many banks are observed over how many years, reaching a number of 4007 institutions. Fortunately, most of the banks have recorded data for all years 2000 – 2007. Note that we excluded several dozen outliers per variable by means of regression deletion diagnostics[15]. This procedure reveals the observations which strongly bias the estimated parameters. We plotted the corresponding density and graphically identified the outliers. After all, we did not globally cut off any percentiles of data or replaced any existing observations by more convenient values (refer to figure 5.5 in the appendix).

We finally accessed SourceOECD Bank Profitability Statistics[16] to verify the coverage of the remaining data in comparison to the total number of savings, cooperative and commercial banks in the respective countries. We report the percentage of covered institutions in table 5.2: Unfortunately, it is obvious that in some countries a serious data reassessment on a Bankscope-based data set finally reveals a rather poor data management on the part of the publisher. Nevertheless, the

[14]Refer to http://cran.r-project.org/web/packages/impute/impute.pdf for details.
[15]So-called 'leave-one-out deletion' (Cook and Weisberg, 1982)
[16]www.sourceoecd.org.

	2000	2001	2002	2003	2004	2005	2006	2007	SUM
CH	147	193	242	256	293	291	294	286	2002
DE	1550	1544	1432	1328	1321	1568	1571	1431	11745
ES	17	17	14	10	65	138	140	87	488
FR	164	172	158	143	150	175	157	115	1234
IT	57	65	45	29	60	572	566	465	1859
SE	7	52	51	52	54	56	59	23	354
GB	30	31	34	37	65	71	69	54	391
US	385	394	374	351	312	303	285	253	2657
SUM	2357	2468	2350	2206	2320	3174	3141	2714	20730

Table 5.1: Observed banks per year and country

absolute number of banks will permit us to estimate cross sectional frontier functions per year and country.

	2000	2001	2002	2003	2004	2005	2006	2007
CH	43.88	59.02	76.58	85.05	97.99	98.64	102.08	100.00
DE	60.59	65.56	65.12	64.43	66.68	81.67	83.70	77.69
ES	6.05	6.05	5.09	3.72	24.44	51.30	51.47	31.07
FR	31.66	33.93	32.85	30.95	40.43	48.48	44.60	34.12
IT	6.78	7.83	5.53	3.68	7.71	72.96	71.37	57.69
SE	5.56	40.31	40.16	41.60	42.86	45.16	47.20	18.70
GB	7.33	8.05	8.95	10.39	18.79	21.19	20.54	16.12
US	4.03	4.24	4.13	3.94	3.58	3.53	3.38	3.05

Table 5.2: Net coverage final bankscope dataset (percent)

5.4.2 Modelling cost frontiers

Descriptive statistics

We refer to the classical Intermediation Approach by Sealey and Lindley (1977) mentioned in chapter 5.2 to set up common and single-equation cost frontiers[17]. We chose to assess cost efficiency rather than

[17] We abstain from a separate assessment of allocative inefficiency vs. technical inefficiency. In that case, additional cost share equations according to Shephard's Lemma have to be estimated.

profit efficiency, as one cannot assume all banks behaving according to profit maximizing considerations. Especially institutions regulated by public law (savings banks in Germany and France) and cooperative banks are usually entrusted with structural or social development tasks. For this reason, cost minimizing rather than profit maximizing behaviour seems to be a common interest across our heterogeneous dataset. Moreover, Bos and Koetter (2007) note that cost functions are problematic if firms incur losses, since the logarithm of non-positive numbers is not defined. Although the authors propose certain procedures to adjust the data and the model, respectively, we would like to avoid the implications of this discussion here.

The descriptive statistics for all variables included in the cost function are given in table 5.3. All currency statements are converted to million USD based on a single current exchange rate. This prevents revaluations of currencies over time to bias our results. Moreover, all nominal values are corrected for inflation with base year 2005.

	mean	median	sd
cost	357.36	39.44	3022.26
loan	4361.00	468.02	29912.10
oea	3790.91	237.65	44390.77
obs	6335.23	52.92	137256.32
ptdep	0.04	0.03	0.23
ppers	0.01	0.01	0.01
ptfa	1.43	0.64	5.30
eqr	7.54	6.17	5.31
com	0.28	0.00	0.45

Table 5.3: Descriptive statistics

We derive the dependent variable *cost* as the sum of personnel, interest and other administrative expenses (Girardone et al., 2009). Furthermore, the output variables *loan* as total loans in million USD, *oea* as other earning assets in million USD and *obs* to account for the growing relevance of off-balance sheet activities in nominal values. The input prices are *ptdep* as price of total deposits (ratio of interest paid to total deposits), *ppers* as price of labour (ratio of personnel expenses to total assets[18]) and *ptfa* as price of total fixed assets[19]. The control variable is *eqr* as equity ratio in percent and serves as risk proxy (the higher the *eqr*, the lower the risk of the bank, e.g. Bos and Schmiedel (2003, p.13)). Additionally, we describe, but do not estimate, the variable *com* as control dummy for commercial institutions[20] in contrast to cooperative and savings banks (the mean reflecting the share of commercial/private institutions in the data set)[21].

Country-specific environmental controls

In chapter 5.2, we discussed the importance of an adequate set of variables explaining country-specific environmental factors. Again, OECD statistical databases deliver comprehensive information on economic, social and regional particularities of all major industrial countries[22].

To assure the availability of a full set of observations in all countries and years, we chose to incorporate three environmental control variables in a common frontier function, in the style of Lozano-Vivas et al. (2002, p.64f.):

- The labour productivity growth rate in percent[23] *prod.growth*. As financial production mainly consists of provision of services, supported by information and communication technology, the productivity growth rate strongly affects banking productivity. We expect labour productivity growth to reduce banking costs.

[18] It is certainly preferable to calculate labour costs as personnel expenses per employee, but the number of employees is a rather poor administered variable in the Bankscope database.

[19] Due to unavailable data on banks' depreciation expenses, we calculated the ratio of other administrative expenses on total fixed assets, see also Girardone et al. (2009).

[20] Note that dummy variables cannot be estimated within fixed effects approaches.

[21] Table 5.9 in the appendix shows the share of commercial, cooperative and savings banks in the respective countries, according to OECD data.

[22] stats.oecd.org.

[23] Labour productivity calculated as gross domestic product per working hour. See OECD for details.

- The gross domestic product per head *gdp.head*. It reflects the wealth and the stage of development of a country. Lozano-Vivas et al. (2002) suggest that a mature environment results in more competitive interest rates and profit margins. Higher interest rates paid typically increase banking costs.

- The number of branches per 1000 inhabitants *branch.dens*. The impact of a denser branch network on banking costs is ambivalent: On the one hand, it facilitates the access to financial services throughout the population; but on the other hand, the capacity of some branches might be underutilized. So Lozano-Vivas et al. (2002) suggest that hight levels of branch density imply high costs and tend to reduce bank efficiency.

Note that, as these variables are not only country-specific, but also time-dependent, they will control for most of the bank-external productivity trends. After having controlled for macroeconomic effects, the remaining 'net' trend parameters we are going to estimate will be reflecting bank-internal productivity changes (managerial abilities, roughly speaking) in a more authentic way.

Functional form

In conformity with the predominant part of the literature, we chose to estimate a translog cost frontier, originally developed by Christensen et al. (1973). With outputs $y_m, m = 1, 2, 3$, input prices $w_k, k = 1, 2, 3$ and control variables $controls_p, p = 1, ..., 4$, (three OECD environmental variables and the risk proxy) given above, the translog cost frontier can be written as[24]:

$$\begin{aligned}\ln cost_{it} &= \beta_0 + \sum_{k=1}^{3}\alpha_k \ln w_{kit} + \sum_{m=1}^{3}\beta_m \ln y_{mit}\\ &+ \frac{1}{2}\sum_{k=1}^{3}\sum_{j=1}^{3}\alpha_{kj}\ln w_{kit}\ln w_{jit} + \frac{1}{2}\sum_{m=1}^{3}\sum_{n=1}^{3}\beta_{mn}\ln y_{mit}\ln y_{nit}\\ &+ \frac{1}{2}\sum_{k=1}^{3}\sum_{m=1}^{3}\gamma_{km}\ln w_{kit}\ln y_{mit} + \sum_{p=1}^{4}\delta_p controls_{pit} + v_{it} + u_{it}\end{aligned}$$

To preserve our notation in the context of (upper) production frontiers, note that symmetry of v_{it} allows us to turn the sign

$$\begin{aligned} \log(cost_{it}) &= \log\left(g(y_{it}, w_{it}); \beta)\right) + v_{it} + u_{it} \qquad |v_{it} \text{ symmetric}\\ \Leftrightarrow -\log(cost_{it}) &= -\log\left(g(y_{it}, w_{it}); \beta)\right) + v_{it} - u_{it}\end{aligned}$$

In words: Mirroring all data in the origin enables us to estimate a negatively shifted intercept by exactly the methods described above, while leaving the slope parameters unchanged.

The existence of 'regular' translog cost and profit frontiers depends on certain regularity conditions (symmetry of the second-order derivatives; linear homogeneity in input prices)[25]:

$$\alpha_{kj} = \alpha_{jk} \quad \text{and} \quad \beta_{mn} = \beta_{nm} \quad \text{for} \quad j, k, m, n = 1, 2, 3$$

and

$$\alpha_1 + \alpha_2 + \alpha_3 = 1$$

in the first order terms, as well as

$$\begin{pmatrix}\alpha_{11} + \alpha_{12} + \alpha_{13} = 0\\ \alpha_{21} + \alpha_{22} + \alpha_{23} = 0\\ \alpha_{31} + \alpha_{32} + \alpha_{33} = 0\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}\gamma_{11} + \gamma_{21} + \gamma_{31} = 0\\ \gamma_{12} + \gamma_{22} + \gamma_{32} = 0\\ \gamma_{13} + \gamma_{32} + \gamma_{33} = 0\end{pmatrix}$$

in the second order terms. With substitution for the third input w_3,

---

[24]Note that inefficiency u_{it} implies a *positive* deviation from the cost frontier.

[25]See e.g., Lang and Welzel (1996).

the restricted translog frontier turns out to be (indices *it* omitted for simplicity)

$$
\begin{aligned}
\ln cost - \ln w_3 = \beta_0 &+ \alpha_1(\ln w_1 - \ln w_3) + \alpha_2(\ln w_2 - \ln w_3)\\
&+ \beta_1 \ln y_1 + \beta_2 \ln y_2 + \beta_3 \ln y_3\\
&+ \alpha_{11}\left(\frac{1}{2}(\ln w_1)^2 - \ln w_1 \ln w_3\right)\\
&+ \alpha_{12}(\ln w_1 \ln w_2 - \ln w_1 \ln w_3 - \ln w_2 \ln w_3)\\
&+ \alpha_{22}\left(\frac{1}{2}(\ln w_2)^2 - \ln w_2 \ln w_3\right)\\
&+ \gamma_{11}(\ln w_1 \ln y_1 - \ln w_3 \ln y_1) + \gamma_{21}(\ln w_2 \ln y_1 - \ln w_3 \ln y_1)\\
&+ \gamma_{12}(\ln w_1 \ln y_2 - \ln w_3 \ln y_2) + \gamma_{22}(\ln w_2 \ln y_2 - \ln w_3 \ln y_2)\\
&+ \gamma_{13}(\ln w_1 \ln y_3 - \ln w_3 \ln y_3) + \gamma_{23}(\ln w_2 \ln y_3 - \ln w_3 \ln y_3)\\
&+ \frac{1}{2}\beta_{11}(\ln y_1)^2 + \beta_{12} \ln y_1 \ln y_2 + \beta_{13} \ln y_1 \ln y_3\\
&+ \frac{1}{2}\beta_{22}(\ln y_2)^2 + \beta_{23} \ln y_2 \ln y_3 + \frac{1}{2}\beta_{33}(\ln y_3)^2\\
&+ \sum_{p=1}^{5} \delta_p controls_p + v + u
\end{aligned}
$$

5.5 Cost efficiency estimation

5.5.1 Discussion of the estimated parameters

We estimated two SFA panel models with deterministic time trend parameters on the basis of the fixed effects extension by Cornwell et al. (1990) (*Identifier*: Fixed effects model **FE**) and Kumbhakar (1990) (*Identifier*: Random effects model **RE**) described above. As we expect all banks in our database to act on an increasingly deregulated common market in the early millenium years, it is virtually meaningless to account for national borders in view of a distinctive international network of financial institutions. For this reason, we can assume identical technology parameters of the cost frontier across all countries, and do not estimate country-specific frontiers. A common frontier goes along with the possibility to set up efficiency rankings, as all banks are assessed against the background of the same benchmark. Nevertheless, all banks in a particular country, whether foreign or domestic banks, face the same environmental conditions which can differ from country to country. In chapter 5.2, we mentioned the recent findings of Berger (2007) who emphasizes the importance of an adequate set of environmental factors controlling for different economic and regional conditions in European and US banking systems. The parameters of the environmental control variables we introduced beside the technology parameters of the cost function adjust country-specific cost frontiers to particular patterns of demand. Moreover, we allow for heterogeneity in the efficiency trends.

A word concerning the optimization procedures: The Maximum-Likelihood optimization for the RE model was implemented within the **R**-Environment (R Development Core Team, 2010). Small simulation studies on an unbalanced sample of 100 firms in eight years (at most) let us expect the estimation of the trend parameters to be unbiased (see figure 5.6 in the appendix; further details on the data generating process are available upon request.). With respect to the FE model, OLS estimation of demeaned panel data is always BLUE.

After optimization we found the parameters given in tables 5.4 and 5.5[26]. In accordance with the literature we observe that the statistical significance of the first- and second-order translog parameters is scattered due to the multicollinearity problem. So we shall focus our discussion on the economic meaning of the remaining parameters.

- The *risk* proxy affects costs in FE und RE in a negative manner. In other words: The higher the equity ratio, i.e. the lower the risk of the bank, the lower the costs. This is somewhat surprising, as we expect equity to be one of the rather cost-intensive sources of bank funding. But apparently, we undervalued the positive signalling of a solid equity funding.

- The parameter of labour productivity growth is the same in both

[26]Note that dummy-variables (e.g., for commercial banking institutions) cannot be incorporated within the estimation, because time-invariant elements disappear in the course of fixed effects regressions.

models: We observe a negative parameter, as expected: Positive productivity growth reduces costs by trend.

- As anticipated by Lozano-Vivas et al. (2002), a higher gross domestic product per head increases banking costs in both models FE and RE.

- Surprisingly, the extra costs of more branches per inhabitant is not definite in both models. In FE, the extra benefits are beyond additional charges, reducing total costs. So we cannot confirm the supposition by Lozano-Vivas et al. (2002). In RE, the corresponding parameter is insignificantly positive.

- Additionally, we compared the mean efficiency gap between commercial and non-commercial (*public*) institutions per country, using the identifier *com*. Figure 5.1 shows the mean efficiency surplus of public banks compared with commercial banks in both models FE and RE. Our result corresponds roughly to the findings of Girardone et al. (2009): The authors note that banks based on mutuality might benefit from local monopoly power.

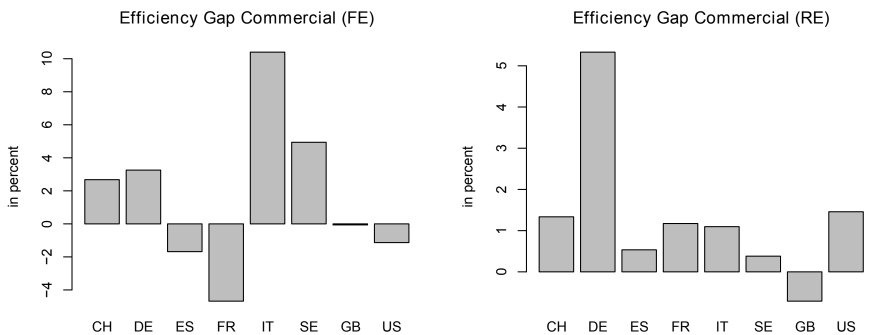

Figure 5.1: Efficiency gap between commercial and non-commercial institutions

Now we turn to the results of the SFA efficiency estimation:

- First, we discuss the time trend parameters ω_1, ω_2 for each country. As we estimated two parameters (linear and quadratic), an (inverse) u-shaped efficiency/intercept change is possible in both settings. Our panel length comprises only eight years at most, so we abstained from the estimation of an additional cubic trend. Figures 5.2 and 5.3 illustrate the country-specific estimated trends with reference to a virtual base year '0'. The resulting efficiency scores and rankings will be discussed in detail below.

  In the case of RE, the sign and value of the parameters can be directly interpreted as *in*efficiency change. In the FE model, the respective parameters describe the changing intercept. A negatively shifted intercept is equivalent to reduced costs. As efficiency is a relative concept in the framework of a fixed effects estimation, no general statement about the resulting efficiency change can be derived from the parameters. We tried to reduce the influence of single outliers on the efficiency scores by setting all banks below the 1% cost quantile as 100% efficient, instead of orientation towards the absolute minimum costs.

- Note that $\hat{\sigma_\tau}$ and $\hat{\sigma_v}$ in the RE model are *not* moments of the estimated values $\hat{\tau}_i$ and $\hat{v}_i$, respectively, but estimators of the parameters of the underlying asymmetric error term distribution $f(\tau) \cdot f(v)$ (cp. section 5.3). Insofar, it is the skewness in the data that determines the estimated values $\hat{\sigma_\tau}$ and $\hat{\sigma_v}$. In terms of the likelihood function optimization: The higher the deviance of the one-sided random variable τ, the more skewed is the distribution of the error term in the frontier function, and the more probable the occurence of high inefficiency observations become.

- In that context, the inefficiency to noise-ratio $\lambda \equiv (\sigma_\tau / \sigma_v)$ indicates how much of the total standard deviation between realized costs and frontier costs in the respective model is due to random noise (good/bad luck) or managerial inefficiency. As we find a value $\lambda \approx 1$, inefficiency and noise are of about equal size in the deviance from the cost frontier. This suggests a moderately skewed error term. (In the extreme case $\lambda \approx 0$, the SFA estimation methods can

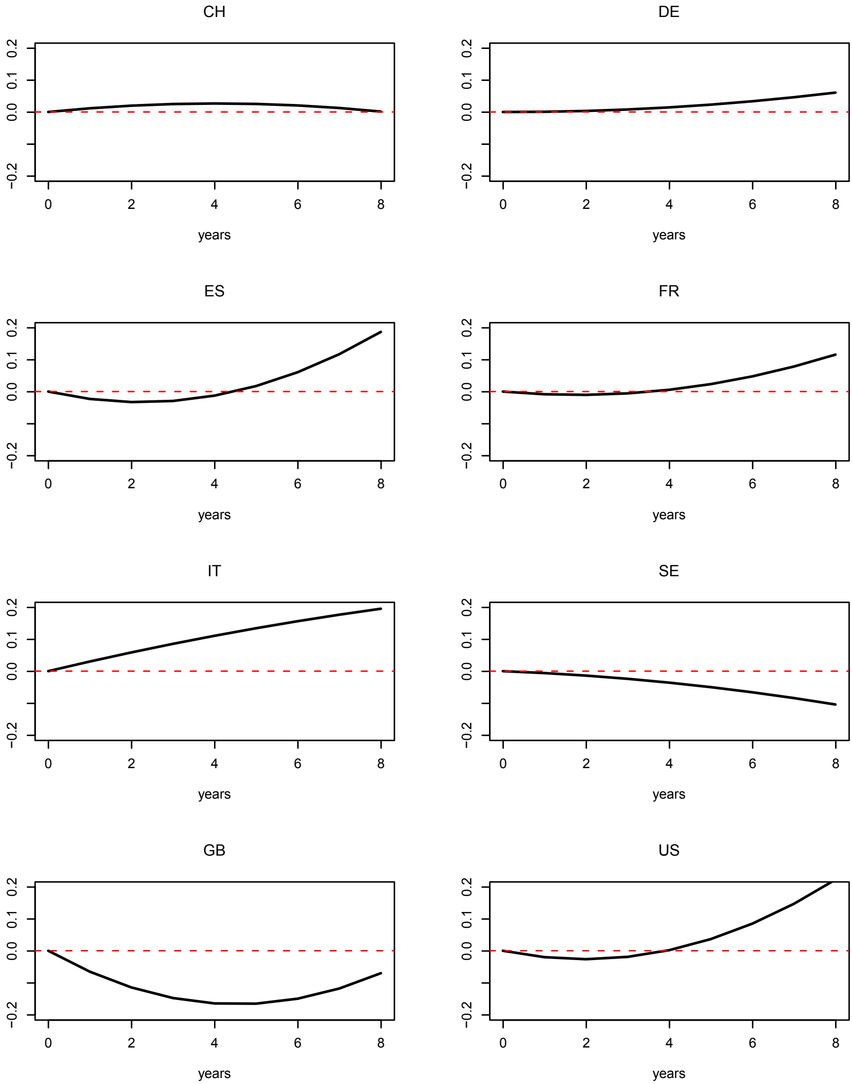

Figure 5.2: Trend of intercept change, fixed effects estimation

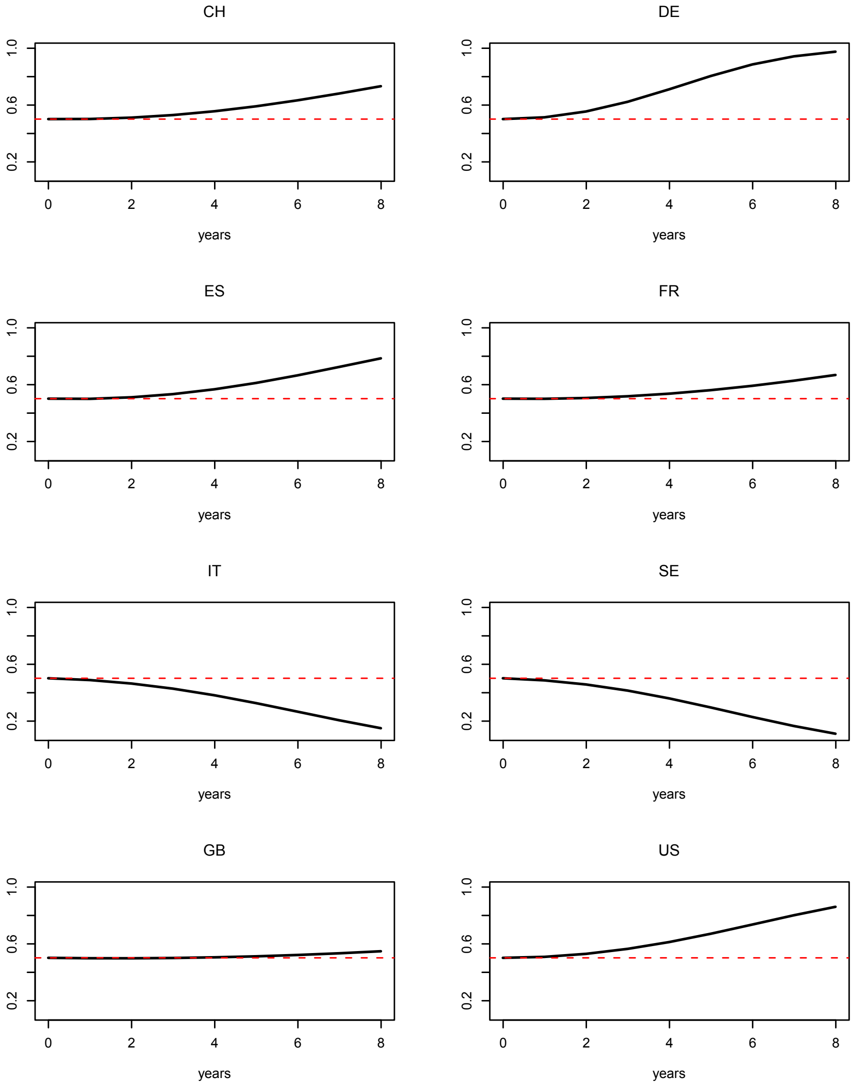

Figure 5.3: Trend of efficiency scores, modified Kumbhakar model

be replaced by OLS estimation, as the error term is not asymmetric at all.)

- Finally, *eps*, the total deviation of the composed error term $\sigma_\varepsilon = \sqrt{\sigma_\tau^2 + \sigma_v^2}$ reflects the total distance between the observations and the frontier function.

5.5.2 Dynamic efficiency ranking

The estimation of a cost frontier enables us to calculate firm-specific efficiency scores according to the procedures discussed in section 5.3. Moreover, as we assumed the deterministic trend parameters to vary from country to country, the firm ranking within a country remains unchanged over the years, but the country's mean efficiency score ranking alternates. Tables 5.6 and 5.7 show the mean efficiency scores and the respective ranking positions per country and year for both models FE and RE. It is obvious that ranks do indeed switch, challenging the popular assumption of a single common rate of efficiency change or even time-invariant efficiency.

Note that the absolute value of the efficiency scores given in the right columns of the tables can only be interpreted against the same benchmark, i.e. within the same model FE or RE. Especially the FE estimation causes the resulting efficiency scores to spread widely.

	Estimate	Std. Error	t value	Pr(>|t|)
a1	0.5353	0.0102	52.4941	0.0000
a2	0.4855	0.0096	50.8161	0.0000
b1	0.4859	0.0106	45.6788	0.0000
b2	0.3971	0.0085	46.9072	0.0000
b3	-0.0179	0.0059	-3.0330	0.0024
a11	0.0810	0.0023	34.6711	0.0000
a12	-0.0664	0.0020	-33.7642	0.0000
a22	0.1055	0.0023	44.9666	0.0000
b11	0.1361	0.0020	67.5890	0.0000
b12	-0.1254	0.0015	-81.9574	0.0000
b13	0.0030	0.0010	3.0758	0.0021
b22	0.1305	0.0015	88.7413	0.0000
b23	-0.0037	0.0009	-4.2196	0.0000
b33	0.0024	0.0008	2.8955	0.0038
g11	0.0242	0.0018	13.6339	0.0000
g21	-0.0099	0.0017	-5.7238	0.0000
g12	-0.0239	0.0012	-19.4894	0.0000
g22	0.0273	0.0013	21.5762	0.0000
g13	0.0080	0.0011	7.0152	0.0000
g23	-0.0166	0.0011	-14.4806	0.0000
risk	-0.0018	0.0003	-5.4816	0.0000
prod.growth	-0.0082	0.0009	-9.5026	0.0000
gdp.head	0.0182	0.0039	4.6576	0.0000
branch.dens	-0.1234	0.0208	-5.9264	0.0000
CH ω_1	0.0131	0.0052	2.5212	0.0117
DE ω_1	-0.0004	0.0030	-0.1252	0.9003
ES ω_1	-0.0297	0.0137	-2.1677	0.0302
FR ω_1	-0.0117	0.0054	-2.1627	0.0306
IT ω_1	0.0309	0.0082	3.7764	0.0002
SE ω_1	-0.0049	0.0120	-0.4115	0.6807
GB ω_1	-0.0736	0.0090	-8.1780	0.0000
US ω_1	-0.0269	0.0043	-6.3209	0.0000
CH ω_2	-0.0016	0.0006	-2.8892	0.0039
DE ω_2	0.0010	0.0003	3.5400	0.0004
ES ω_2	0.0066	0.0011	5.8335	0.0000
FR ω_2	0.0033	0.0006	5.5543	0.0000
IT ω_2	-0.0008	0.0006	-1.2948	0.1954
SE ω_2	-0.0010	0.0013	-0.7883	0.4305
GB ω_2	0.0081	0.0009	8.7560	0.0000
US ω_2	0.0068	0.0004	17.0022	0.0000

Table 5.4: Regression Results Fixed Effects

	Estimate	Std. Error	t value	Pr(>|t|)
intercept	1.7022	0.0493	34.5303	0.0000
a1	0.7182	0.0086	83.5802	0.0000
a2	0.3216	0.0182	17.6730	0.0000
b1	0.5109	0.0022	232.4112	0.0000
b2	0.4779	0.0013	370.2710	0.0000
b3	-0.0121	0.0054	-2.2412	0.0250
a11	0.0468	0.0019	25.0030	0.0000
a12	-0.0273	0.0018	-15.5417	0.0000
a22	0.0469	0.0044	10.6459	0.0000
b11	0.1607	0.0024	66.2222	0.0000
b12	-0.1538	0.0023	-65.5311	0.0000
b13	-0.0034	0.0015	-2.2500	0.0245
b22	0.1499	0.0006	242.1546	0.0000
b23	0.0020	0.0009	2.2179	0.0266
b33	0.0002	0.0024	0.0963	0.9233
g11	-0.0010	0.0051	-0.1853	0.8530
g21	0.0152	0.0052	2.9402	0.0033
g12	-0.0345	0.0053	-6.5773	0.0000
g22	0.0381	0.0037	10.2616	0.0000
g13	0.0229	0.0012	19.7350	0.0000
g23	-0.0372	0.0014	-27.0145	0.0000
risk	-0.0048	0.0002	-19.9662	0.0000
prod.growth	-0.0103	0.0013	-8.0112	0.0000
gdp.head	0.0036	0.0003	12.7804	0.0000
branch.dens	0.0201	0.0270	0.7439	0.4570
CH ω_1	0.0145	0.0678	0.2136	0.8308
DE ω_1	0.0083	0.0302	0.2752	0.7832
ES ω_1	0.0280	0.1440	0.1943	0.8460
FR ω_1	0.0160	0.0585	0.2739	0.7842
IT ω_1	0.0261	0.0084	3.1067	0.0019
SE ω_1	0.0299	0.2379	0.1257	0.9000
GB ω_1	0.0159	0.0585	0.2727	0.7851
US ω_1	0.0003	0.0690	0.0044	0.9965
CH ω_2	-0.0175	0.0095	-1.8294	0.0674
DE ω_2	-0.0579			
ES ω_2	-0.0236	0.0189	-1.2504	0.2112
FR ω_2	-0.0128	0.0092	-1.3955	0.1629
IT ω_2	0.0240			
SE ω_2	0.0290	0.0856	0.3384	0.7351
GB ω_2	-0.0049	0.0090	-0.5468	0.5845
US ω_2	-0.0283	0.0096	-2.9542	0.0031
sigma.noise	0.1227	0.0014	90.4152	0.0000
sigma.tau	0.1425	0.0014	98.9262	0.0000
lambda	1.1618			
eps	0.1881			

Table 5.5: Regression Results, Modified Kumbhakar-Model

	2000	2001	2002	2003	2004	2005	2006	2007
mean CH	0.6519	0.6629	0.6708	0.6838	0.7014	0.7232	0.7461	0.7674
mean DE	0.5890	0.6023	0.6098	0.6187	0.6283	0.6381	0.6452	0.6471
mean ES	0.5209	0.5393	0.5466	0.5489	0.5457	0.5364	0.5191	0.4926
mean FR	0.5860	0.6015	0.6087	0.6146	0.6183	0.6195	0.6151	0.6034
mean IT	0.7522	0.7497	0.7424	0.7394	0.7397	0.7427	0.7450	0.7441
mean SE	0.6825	0.7048	0.7236	0.7474	0.7758	0.8086	0.8421	0.8723
mean GB	0.5245	0.5650	0.5939	0.6167	0.6319	0.6384	0.6333	0.6146
mean US	0.7440	0.7664	0.7734	0.7734	0.7657	0.7498	0.7228	0.6834
rank CH	4.0	4.0	4.0	4.0	4.0	4.0	2.0	2.0
rank DE	5.0	5.0	5.0	5.0	6.0	6.0	5.0	5.0
rank ES	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
rank FR	6.0	6.0	6.0	7.0	7.0	7.0	7.0	7.0
rank IT	1.0	2.0	2.0	3.0	3.0	3.0	3.0	3.0
rank SE	3.0	3.0	3.0	2.0	1.0	1.0	1.0	1.0
rank GB	7.0	7.0	7.0	6.0	5.0	5.0	6.0	6.0
rank US	2.0	1.0	1.0	1.0	2.0	2.0	4.0	4.0

Table 5.6: Efficiency Ranking FE

	2000	2001	2002	2003	2004	2005	2006	2007
mean CH	0.9179	0.9165	0.9137	0.9097	0.9044	0.8982	0.8912	0.8837
mean DE	0.9447	0.9404	0.9335	0.9246	0.9153	0.9074	0.9019	0.8987
mean ES	0.9248	0.9233	0.9201	0.9153	0.9090	0.9015	0.8934	0.8852
mean FR	0.9244	0.9237	0.9220	0.9194	0.9160	0.9117	0.9068	0.9013
mean IT	0.9512	0.9535	0.9571	0.9616	0.9671	0.9731	0.9791	0.9848
mean SE	0.9419	0.9452	0.9502	0.9567	0.9643	0.9722	0.9799	0.9865
mean GB	0.8803	0.8803	0.8799	0.8789	0.8774	0.8754	0.8730	0.8700
mean US	0.9323	0.9296	0.9253	0.9194	0.9124	0.9046	0.8970	0.8901
rank CH	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0
rank DE	2.0	3.0	3.0	3.0	4.0	4.0	4.0	4.0
rank ES	5.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0
rank FR	6.0	5.0	5.0	5.0	3.0	3.0	3.0	3.0
rank IT	1.0	1.0	1.0	1.0	1.0	1.0	2.0	2.0
rank SE	3.0	2.0	2.0	2.0	2.0	2.0	1.0	1.0
rank GB	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
rank US	4.0	4.0	4.0	4.0	5.0	5.0	5.0	5.0

Table 5.7: Efficiency Ranking RE

Discussion of the FE ranking

According to the FE model we identify the most efficient banking systems in Sweden, Italy and the USA, and the most inefficient in Spain (rank eight in all years). The mean efficiency score gap between best and worst countries in all years is about 35%. Italian banks lose their first ranking position in 2000 to the Swiss and the Swedish banking systems and end in the third rank. This seems to be due to strictly increasing costs, compared to the other countries (table 5.6 on page 134). As Sweden is the only country exhibiting stictly decreasing costs, the ranking position clearly ameliorates and ends in the first rank in 2007. Germany and France hold moderate ranking positions. Especially the distinctive u-shaped cost trend in GB strongly reflects the corresponding ranking positions.

Discussion of the RE ranking

We now turn to the discussion of the ranking and efficiency scores according to the RE model. As Sweden and Italy are the countries with the strongest, strictly monotonic decreasing inefficiency (table 5.7 on the previous page), the relative ranking positions level at first and second position, respectively. The mean efficiency scores between both countries do not differ much.

Except for GB (with insignificantly estimated trend parameters $\omega_1 = \omega_2 \approx 0$), all other countries exhibit increasing inefficiency, at most in Germany and the USA. Consequently, both countries lose ranking positions against France (with a rather low increase in inefficiency). The worst ratings of GB banking systems (rank eight) and the Swiss banking system (rank seven) do not alter over time.

In order to guard against misunderstandings we must underline that we find only small differences in mean efficiency scores across countries and years in the RE model. Actually, the range is no more than ten percentage points from the most efficient to the most unefficient banking system. Together with overall high efficiency values of about 90%, this may allow the conclusion that all observations across the countries are very close to the common frontier function. Our assumption of a common European and US banking technology therefore seems to be justified.

5.5.3 A note on the inter-model comparison

We presented the results of two different approaches. It is not surprising that the estimated efficiency scores differ. In this regard, the prevailing opinion in the literature agrees that inter-model comparisons of efficiency scores are a rather problematic undertaking. But there is nevertheless the slight hope that efficiency rankings in different approaches are more or less consistent[27]. With a view to tables 5.6 and 5.7 on page 135 we actually find some coherence between the FE and the RE rankings, as well as the trend parameters. Although the assessments of the Swiss and the US banking systems strongly differ, there is common evidence that the British and Spanish banks perform rather poorly, whereas the Swedish and Italian banks seem to be fairly cost efficient.

Now the question might arise which model is to be preferred? What we can say is that the semiparametric Fixed Effects model rests on less restrictive assumptions: In particular, no decision concering the distribution of the inefficiency term has to be made, and correlation between inefficiency and explanatory variables is permitted. So, as the assumption of Random Effects might be violated, it shall be safer to rely on the Fixed Effects estimation. However, efficiency estimates in FE-models are known to be strongly dependent on outliers. We tried to reduce this dependence using the 1% percentile as a benchmark, rather than a single 'outlier'.

The application of two different methods in the course of our empirical analysis resulted in similar findings in efficiency trends and rankings, rendering our results more credible. For reference, we also quote similar results derived by Girardone et al. (2009), Bos and Schmiedel (2003) (table 5.10 on page 141 in the appendix).

[27] As the fundamental idea to any SFA method is the determination of an intercept shift of the average cost/production function, given a set of slope parameters identical or very similar to OLS, the vertical position of each oberservation relative to each other is unchanged.

5.6 Conclusion

In this study we estimated a cost frontier in accordance with the standard procedure we presented in the literature overview (chapter 5.2). We put forward an alternative reasoning concerning the degree of heterogeneity in international bank production/cost functions. The prevailing approach to cross-country efficiency comparisons consists either in estimation of a common frontier without heterogeneity in efficiency trends or in the estimation of country-specific frontiers accounting for perfect heterogeneity in trends and technology. In the latter case, efficiency differences among countries cannot be revealed and one loses the possibility to rank the countries' banking systems according to the background of a common technological benchmark. Nevertheless, there is clear agreement in the literature that efficiency rankings are more informative than absolute efficiency scores.

We showed that a compromise between either approach is possible. Beyond that, the results we obtained advise us to always consider specific efficiency trends per country: Obviously, efficiency rankings are not as static as often assumed, especially in consideration of the fact that panel lengths of 5 years or more are usual in bank efficiency studies (cp. table 5.8).

Basically, non-parametrical efficiency trends can also be derived by successive estimation of cross-sectional SFA models[28]. But we mentioned that the resulting estimators $\hat{u}_i$ are inconsistent and should be avoided in favour of panel data methods. Admittedly, the implementation of the modified Kumbhakar (1990) model allowing for country-specific efficiency trends is cumbersome in the course of empirical analyses, because to our knowledge, no publicly available software implementation exists. The robust alternative is the simple fixed effects approach on the basis of Cornwell et al. (1990).

[28] Or the successive application of DEA per year, cp. Casu and Girardone (2006).

5.7 Appendix

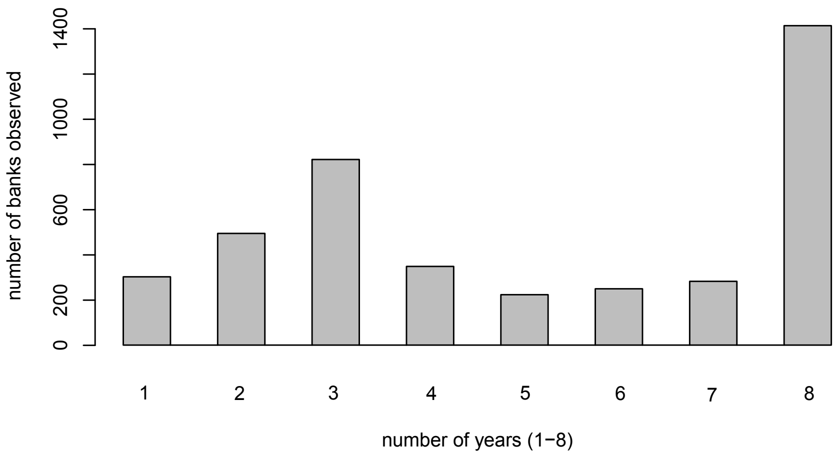

Figure 5.4: Number of observed banks, per years

Study	Region covered	Years	Method	Frontier
Al-Sharkas et al. (2008)	CALL Report USA	1986 – 2002	SFA, DEA	Cost, Profit
Altunbas et al. (2001)	EU 15	1989 – 1997	SFA	Cost
Bos and Schmiedel (2003)	BEl, FRA, GER, ITA, NED, ESP, SUI, GBR	1993 – 2000	SFA Meta Frontier	Cost, Profit
Casu and Molyneux (2003)	FRA, GER, ITA, ESP, GBR	1993 – 1997	DEA (VRS)	Production
Casu and Girardone (2006)	EU 15	1997 – 2003	DEA (VRS)	Production
Dietsch and Lozano-Vivas (2000)	FRA, ESP	1988 – 1992	DFA	Cost
Girardone et al. (2009)	EU 15	1998 – 2003	SFA	Cost
Goddard et al. (2004)	DEN, FRA, GER, ITA, ESP, GBR	1992 – 1998 (balanced)	OLS	Profitability
Lozano-Vivas et al. (2002)	BEL, DEN, FRA, GER, ITA, LUX, NED, POR, ESP, GBR	1993	DEA (SFA)	Production
Oliveira and Tabak (2005)	41 Countries worldwide	1995 – 2002	DEA (VRS)	Profitability
Pastor et al. (1997)	USA, AUT, ESP, GER, GBR, ITA, BEL, FRA	1992	DEA (CRS, VRS)	Production
Pastor and Serrano (2006)	AUT, GER, BEL, DEN, ESP, FRA, GRE, ITA	1992 – 1998	DEA	Production
Weill (2004)	FRA, GER, ITA, ESP, SUI	1992 – 1998 (balanced)	SFA, DEA, DFA	Cost

Table 5.8: Overview of selected bank efficiency studies

	commercial	savings	cooperative
CH	71.72	27.95	0.33
DE	8.50	23.54	67.96
ES	51.76	17.15	31.10
FR	69.22	0.00	30.78
IT	38.64	0.00	61.36
SE	41.39	58.61	0.00
GB	100.00	0.00	0.00
US	88.26	10.93	0.81

Table 5.9: commercial/public banking relation per country

	Girardone et al. (2009)		Bos and Schmiedel (2003)	
	Mean CE	Rank	Mean Metafrontier CE	Rank
CH	NA	NA	87.95%	3
DE	77.53%	5	80.41%	6
ES	78.10%	3	89.38%	2
FR	79.79%	2	85.01%	4
IT	89.12%	1	95.19%	1
SE	77.79%	4	NA	NA
GB	67.41%	6	83.05%	5
US	NA	NA	NA	NA

Table 5.10: Similar static ranking results

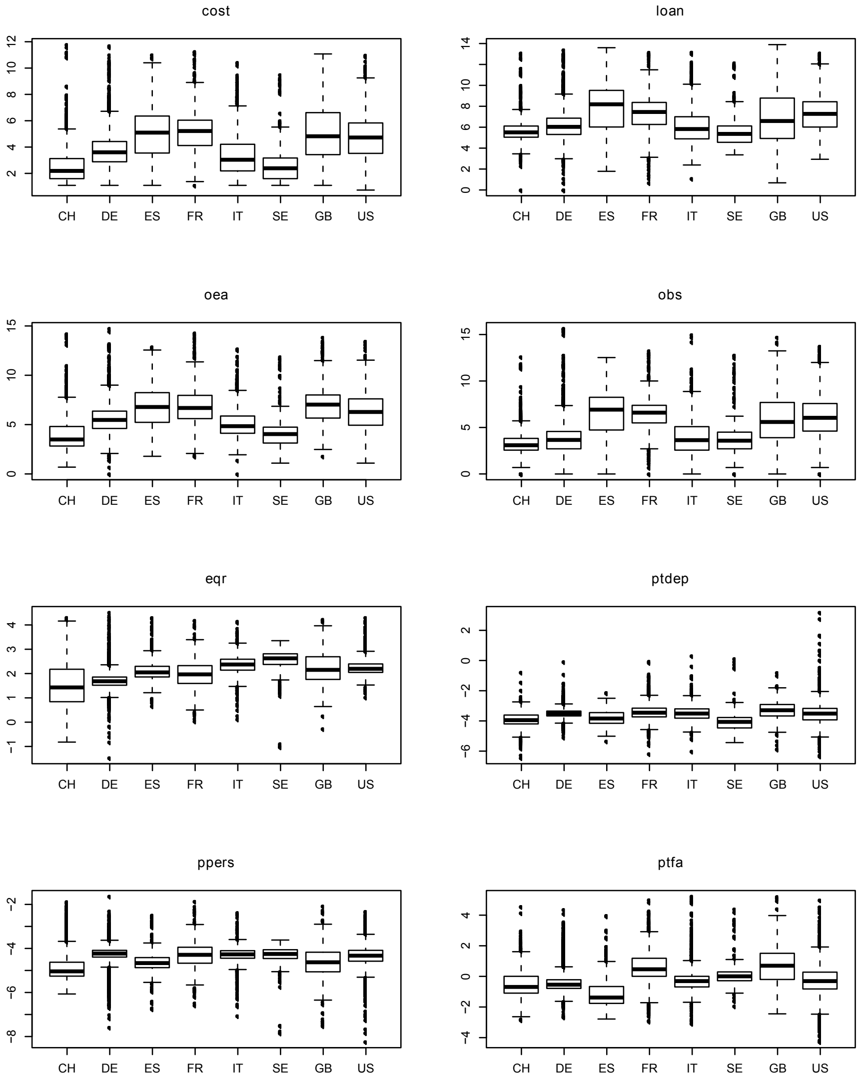

Figure 5.5: Boxplots outputs and input prices, in logs

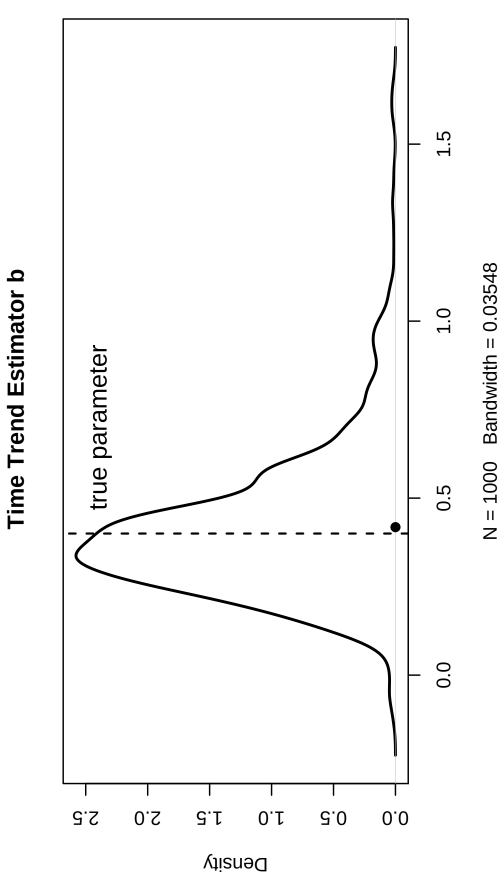

Figure 5.6: Simulated distribution of a single time trend parameter

Chapter 6

6 Bank efficiency estimation based on quantile cost functions with fixed effects

6.1 Introduction

Driven by the outstanding importance of financial institutions for economies worldwide, the assessment of national banking systems is in the vital interest of financial policy makers, banking supervisors (authorities) and economic science. The resulting glut of studies revealing different determinants of efficient banking activities is rather intransparent for occasional readers. It is the merit of authors like Berger et al. (1993) who early summarized the key messages of the studies hitherto published. Over a decade later, Berger (2007) again outlined the recent developments in measuring bank efficieny against the background of common international financial markets.

Bank efficiency estimation basically consists of the modelling of production, profit or cost functions of financial firms, using balance sheet and income statement data. Different approaches exist to fit not only the conditional mean function but an 'efficient frontier' to the data. Among the most popular approaches are the parametric Stochastic Frontier Analysis SFA (Aigner et al., 1977, Meeusen and van de Broeck, 1977) and the non-stochastic, non-parametric Data Envelopment Analysis DEA (Charnes et al., 1978). The respective methodologies have developed fast in recent decades, and we have always observed empirical bank efficiency studies soon applying recently introduced methodological improvements. Moreover, even studies focused on efficiency methods often used banking data to illustrate new approaches (Greene, 2005a, 2002). So it is our supposition that the needs of applied bank efficiency estimation and literature on efficiency frontier methods have always enriched each other mutually. We suppose this significantly accelerated the development in both fields.

In the progress of this study, we are going to pursue the discussion on the applicability of Quantile Regression QR (Koenker and Bassett, 1978) in the course of efficiency estimations. Basically, Quantile Regression focuses on the conditional quantiles of the dependent variable given a set of explanatory variables. Insofar, it characterizes the whole conditional distribution of the dependent variable without any assumptions regarding the random noise. Consequently, QR can be seen as a generalization of linear models aiming only at the conditional mean of the dependent variable, thereby imposing certain restrictions on the random noise term (e.g. Ordinary Least Squares).

As all efficiency methods seek to estimate an efficient frontier enveloping the 'data cloud', it is only natural to consider high and low conditional quantiles as frontier functions. The only approaches to bank-specific cost functions we found are those of Guala (2008) and Behr (2010). Both studies are based on cross-sectional data, estimating low quantile cost functions that constitute the efficient cost frontier. The authors note that resulting frontier cost function parameters considerably differ from shifted mean regressions like SFA. If available, it is common practise in efficiency estimation to make use of repeated observations, i.e. panel data[1]. Quantile Regression for panel data has been introduced by Koenker (2004), but was only rarely applied in economics. We only found two applications to problems of labour economics, e.g. the estimation of the Mincer Equation (Bargain and Kwenda, 2009, Kniesner et al., 2009). To our best knowledge, Quantile Regression with panel data has never been used for the purpose of efficiency estimation. So we are about to explore the new opportunities it opens. In particular, we will put forward an 'efficiency sector concept', providing robust information on time-variant firm-specific efficiency trends.

The weak extension of the suggested methods in empirical literature may be traced back to the fact that optimization algorithms are not yet

[1] The DEA is the only efficiency method that does not make use of repeated observations.

implemented in standard statistical software. The solution to Quantile Regression problems with longitudinal data is not straighforward, even though Roger Koenker provides some kind of 'prototype function' to users of the R Environment. Moreover, Koenker himself treats Quantile Regression with longitudinal data as methodologically not mature: In his comprehensive monograph (Koenker, 2005) we find this topic in the chapter 'Twilight Zone of Quantile Regression'. So basically, we are now facing two challenging tasks: First to estimate and interpret panel data Quantile Regression parameters, and second to use the estimated parameters in the course of an efficiency analysis. For illustrative purposes, we perform an application estimating bank efficiency in German and US commercial banks for the years 2000 – 2007.

Although just a first attempt, we expect to see our study encouraging a new discussion on the use of alternative efficiency methods like Quantile Regression. Ideally, this would again lead to a mutual enrichment in both fields. For these purposes the remainder of this study is composed as follows: In chapter 6.2, we corrobarate efficiency estimation on the basis of Quantile Regression with fixed effects by the recapitulation of both components: Efficiency estimation in conditional mean models with fixed effects (section 6.2.1) and efficiency estimation in cross-sectional conditional quantile models (6.2.2). Our new proposition is described in sections 6.2.2 and 6.2.2. In chapter 6.3 we discuss the results of an exemplifying application, using cost function data on German und US banks. Chapter 6.4 finally contains the conclusion.

6.2 Methodology

In this chapter we lay down the basic principles of efficiency measurement on the basis of firm-specific fixed effects estimations. Whereas the traditional fixed effects approach in panel data analyses is a standard procedure, the cost frontier interpretation of a lower boundary constituted by the minimum firm-specific intercept is rather unusual in empirical literature. Nevertheless, this approach has already been discussed a long time ago (Schmidt and Sickles, 1984) and has certain advantages over Stochastic Frontier Analyses (Meeusen and van de Broeck, 1977, Aigner et al., 1977) based on the a-priori assumption of a specific parametric inefficiency distribution.

6.2.1 Linear regression cost frontiers with fixed effects

The basic idea of efficiency estimation with panel data is described in Schmidt and Sickles (1984, p. 368 f.). A log-linear cost function for firms $i = 1, ..., N$ in times $t = 1, ..., T$ can be written as:

$$c_{it} = \alpha + \sum_{k=1}^{K} \beta_k x_{kit} + v_{it} + u_i \tag{6.1}$$

with c_{it} costs, α common intercept parameter (i.e. minimum possible fixed costs), x_{kit} the $k = 1, ..., K$ time-varying explanatory variables without a constant, $v_{it} \sim iid.N(0, \sigma_v)$ random noise and $u_i \geq 0$ firm-specific and time-invariant inefficiency without any parametric distributional assumptions[2]. The v_{it} are uncorrelated with x_{kit} and u_i, but the u_i may be correlated with the regressors.

As it is rather difficult to decompose the estimation of α and u_i due to perfect multicollinearity, we simply assume that the minimum u_i equals zero. I.e., the most cost efficient firm in the sample has fixed costs of exactly α and inefficiency u_i of zero. This assumption is asymptotically justified in the case $N \to \infty$. As a result, we define the firm-specific amount of inefficiency as vertical distance between a true but unobserved α constituting the efficient lower boundary and the firm specific effect α_i:

$$\alpha_i \equiv \alpha + u_i \tag{6.2}$$

This leads to the classical fixed effects model

$$c_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k x_{kit} + v_{it} \tag{6.3}$$

which can be estimated by either dummy variables indicating each i or by groupwise demeaning (or first differencing) to 'get rid of' the α_i.

[2] At least, $E(u_i)$ and $Var(u_i)$ exist.

The estimated $\hat{\alpha}_i$ are consistent when $T \to \infty$[3]. Efficiency scores CE_i can be recovered in the case of log-linear Cobb-Douglas cost functions as:

$$\widehat{CE}_i = \left(\frac{\exp(\hat{\alpha}_i)}{\exp(\hat{\alpha})} \right)^{-1} \in [0, 1] \tag{6.4}$$

with $\hat{\alpha} = \min(\hat{\alpha}_i)$.

Obtaining information on the inference of the $\hat{u}_i$ is not straighforward. As typically $N \gg T$ but far from ∞, the `min` operation considerably matters. Upon certain terms Schmidt and Sickles (1984) suggest that the assumption of a double-exponential distribution of the u_i facilitates evidence on standard errors and confidence intervals. But normally, there is no way around bootstrapping procedures[4].

Although uncomplicated and robust in estimation, the fixed effects approach is not above criticism. As several authors note, some serious drawbacks exist: Greene (2005a) gives us a warning of 'superficially' treating the fixed effects approach to inefficiency as a 'trivial extension' of the basic stochastic frontier model: '[...] the model is not a simple reparameterization, it is a substantive reinterpretation of the model components'[5]. Kumbhakar and Lovell (2003, p. 100) substantiate: The fixed effects do not only capture technical efficiency, but also the effects of all environmental phenomena. So if in fact T is fixed, Cornwell and Schmidt (2008) expect efficiency scores to be biased downward due to the `min` operation.

Even in the (hypothetical) case of a full specification of all environ-

[3] Kumbhakar and Lovell (2003, p. 107) state: 'The longer the panel, the less likely it becomes that technology remains constant.' Given the statement above the contradiction is understandable: The longer the time period in the panel, the better the estimator $\hat{u}_i$, but the less tenable the assumption of time independence becomes (Greene, 2008, Cornwell and Schmidt, 2008). Against that background, the statement of consistency of a time-invariant u_i in the case $T \to \infty$ is an economic antagonism.

[4] We observe the same problem in the Distribution Free Approach (DFA) introduced by Berger (1993). Here too, efficiency scores are calculated in comparison to the 'best-practice' firm in the sample.

[5] Consequently, Greene puts forward a 'true fixed effects' approach, not mixing up inefficiency u_i and firm-specific effect α_i.

mental variables, one might lose some other characteristics of the firm due to a failure in the 'within optimization' (Greene, 2008): In the presence of time-invariant attributes of the firm, e.g. classification dummies, the individual-mean correction leads to multiple observed zeros for the firm. Omitting these effects from the model will cause them to reappear in the fixed effects, masked as (in)efficiency.

6.2.2 Quantile regression cost frontiers

The cross-sectional case

Quantile regession is based upon the pioneering work of Koenker and Bassett (1978). The authors introduced an optimization strategy to define the median and other quantiles of sample observations as a solution to a minimization problem of (a)symmetrically weighted absolute residuals. The extension to *conditional* quantiles was obvious and soon constituted an alternative to conditional mean regressions. A comprehensive monograph also covering recent developments in Quantile Regression can be found in Koenker (2005), and an accessible introduction with details on optimization procedure and inference in Hao and Naiman (2007).

Let us consider the classical least squares optimization rule with c_i again the endogeneous variable and x_i a $(K+1) \times 1$-dimensional vector of exogeneous variables, including a constant:

$$\min_{\beta \in \mathcal{R}^K} \sum_i (c_i - \mu(x_i, \beta))^2$$

This delivers the conditional mean expectation function $E(c|x) = \mu(x'\beta)$. Otherwise, the minimization of absolute residuals leads to:

$$\min_{\beta \in \mathcal{R}^K} \sum_i |c_i - \zeta(x_i, \beta)|$$

This is the conditional median function $M(c|x) = \zeta(x'\beta)$ (Hao and Naiman, 2007, p.34). In general, any conditional τ-quantile function $Q_\tau(c|x) = \zeta(x'\beta(\tau))$ with $\tau \in (0,1)$ can be estimated by

$$\min_{\beta \in \mathcal{R}^K} \sum_i \rho_\tau(c_i - \zeta(x_i, \beta)) \tag{6.5}$$

The so-called *check*-function $\rho_\tau(z)$ is

$$\rho_\tau(z) = \begin{cases} \tau \cdot z & \text{if} \quad z > 0 \\ (\tau - 1) \cdot z & \text{if} \quad z < 0 \end{cases} \tag{6.6}$$

Weighting positive deviations from the target value ζ with τ and negative deviations with $\tau - 1$ replaces the conventional calculation of quantiles by means of sorting and counting. So we are facing a linear optimization problem which can be solved by Simplex methods described in detail in Koenker (2005, Chapter 6). The procedures are available within most statistical computer programs.

As Koenker and Hallock (2000) state, the advantages of Quantile Regressions compared to traditional mean expectation functions are obvious. Just as we are not always pleased with information on sample means, we would like to gain deeper insight into the whole sample distribution. Information on every quantile of the sample distribution allows us to calculate alternative measures on location (e.g. median) and shape (e.g. skewness) (Hao and Naiman, 2007, p.12 ff.).

Moreover, in Quantile Regression we can abstain from any assumption concerning the random noise. In particular, the assumption of *iid.* Gaussian errors is rather restrictive. General least squares regression in the case of heteroscedastic errors wipes out any further information. In the same way, asymmetrically skewed disturbances remain undetected when focusing on conditional means. Only Quantile Regression reveals the characteristics of the noise distribution without the need to make any a-priori assumptions.

The idea of efficiency measurement based on Quantile Regression for cross-sectional data is described in Behr (2010). He and other authors before (Bernini et al., 2004, Liu et al., 2008) noticed that frontier production/cost function parameters (input-/output-elasticities) at high/low conditional quantiles may strongly differ from the conditional mean estimates we find in SFA analyses[6], for instance.

To our knowledge, Liu et al. (2008) is the first study interpreting negative deviations from a $\tau = 0.8$ production frontier as inefficiency. And Behr (2010) is the first bank efficiency study not only estimating quantile cost functions for multiple $0 < \tau < 1$, but also interpreting positive deviations in costs relative to the best practice cost frontier as inefficieny. In particular, he assumes a $\tau = 0.05$ cost function to represent the firms' common benchmark. All observations lying below the cost frontier are fully efficient, whereas observations above the frontier are inefficient. The vertical distance has the same meaning as u_i in equation (6.1).

In comparison to SFA analyses, one might ask where we can find the distinction between inefficiency and random noise. In fact, this approach does not differentiate between good/bad luck and inefficiency. Normally, this is not necessarily a problem, as non-econometric approaches to efficiency measurement, e.g. the Data Envelopment Analysis (DEA), basically do not account for random noise, either. The typical consequence we observe in non-stochastic approaches is the fact that inefficiency scores usually tend to be larger than in Stochastic Frontier Analyses. This observation is confirmed in every DEA-SFA comparison (Al-Sharkas et al., 2008, Hjalmarsson et al., 1996, Bauer et al., 1998, Weill, 2004), as well as in Behr (2010) for the Quantile Regression.

What we consider more critically is any efficiency measurement based on cross-sectional data. Not only cross-sectional SFA leads to inconsistent estimates of firm-specific inefficiency (Kumbhakar and Lovell, 2003, p.78), cross-sectional observations basically lack the chance to verify the data consistency with respect to firm-specific 'unusual' observations. Only longitudinal data provide multiple observations on every firm to ensure estimation of consistent efficiency scores, as we have seen above. The extension of Quantile Regressions to panel data and the resulting new opportunities for efficiency measurement will be discussed in the next section.

[6]In simple terms, a stochastic frontier is just a vertically shifted mean regression line.

Panel data structure with fixed effects

After having presented two completely different approaches to efficiency analyses, it is about time to consider the potential benefits a conditional Quantile Regression with longitudinal data may bring.

Koenker (2004), Lamarche (2006) propose the following model with firm-specific fixed effects for all $i = 1, ..., N$:

$$Q_{c_{it}}(\tau | X = x_{it}) = \alpha_i + x'_{it}\beta(\tau) \tag{6.7}$$

which can be solved by minimizing (for the ease of notation let $T_i = T$ for all i)

$$\min_{(\beta,\alpha)\in\mathcal{R}^{(K+1)+N}} \sum_{q=1}^{Q}\sum_{i=1}^{N}\sum_{t=1}^{T} w_q \rho_{\tau_q}(c_{it} - \alpha_i - x'_{it}\beta(\tau_q)) \tag{6.8}$$

with ρ_{τ_q} the usual check function given in equation (6.6). All other variables with the exception of w_q are known. It is important to note that the x_{it} contain a constant. Obviously, Koenker suggests optimizing several quantiles τ_q, $q = 1, ..., Q$ simultaneously. The corresponding w_q does not adjust the relative weight of the respective Quantile Regression on the estimation of the $\beta_k(\tau_q)$, but on the estimation of the effects α_i. This facilitates the disentanglement of multiple intercepts specified in model (6.8), as we shall soon see.

In particular, we are to estimate firm-specific intercepts α_i as well as Quantile Regression intercepts $\beta_0(\tau_q)$. Koenker himself speaks of an individual 'location shift' of the whole conditional distribution $c|x$. But thinking it over in more detail, especially when considering a τ-specific $\alpha_i(\tau)$, he admits that the differenciation may actually 'strain credulity'. Canay (2010), for example, notes that Quantile Regression already allows the researcher to account for unobserved heterogeneity. So what do the fixed effects stand for?

Kniesner et al. (2009) are on their way to a tentative possible explanation: Quantile Regression with fixed effects allows the researcher to disentangle two kinds of heterogeneity in the data: First, the latent person-specific heterogeneity α_i that we shall call 'econometric (technical, due to omitted variables) heterogeneity', and second, the 'economic heterogeneity' represented by the $\beta_0(\tau)$.

So it seems evidently advantageous to collect as much information as possible by estimation of multiple Quantile Regression lines simultaneously to disentangle the heterogeneity in the model: The estimation of single Quantile Regressions one after the other would most likely lead to non-conformity of the α_i in each case, as we are encountering an identification problem. Moreover, performing bootstrapping procedures (Kato et al., 2010) on all parameters simultaneously delivers an estimated variance-covariance matrix that allows the researcher to test for inter-quantile differences of the respective $\beta_k(\tau_q)$ (e.g. Wald-test, refer to Hao and Naiman (2007, p.43 f.)).

Unfortunately, the optimization of the model is not straightforward. Especially when N is large, the number of parameters α_i tends to exhaust computer capacity: As standard demeaning techniques cannot be applied[7], dummy matrices contain the index information on the α_i's in N columns. As most entries of these matrices are zero, Koenker and Ng (2005) discuss 'sparse' linear algebra methods to solve the corresponding linear programming problems. The procedures were made available for the R environment (R Development Core Team, 2010) by the authors.

The contribution to efficiency estimation

Now the central question that arises is: Is the model a contribution to efficiency estimation methods? We put forward our opinion that the model given in equation (6.8) does not contribute to efficiency estimation. In particular, we have to ask what the benefit of (6.8) in comparison to conditional mean fixed effects (6.3) is. As Canay (2010) notes, the α_i in both models are the same, as we are talking about the same distribution $c_{it}|x_{it}, \alpha_i$. So the efficiency interpretation of the firm-specific fixed effects shares the same shortcomings discussed above and does not lead to any new insight. Moreover, when (in-)efficiency is captured by the heterogeneity in α_i, what heterogeneity is represented by different τ-quantiles? Obviously, the best-practice interpretation of lower cost quantiles does not hold in this case.

[7] Canay (2010) alternatively suggests a two-step estimator with first differencing.

So we suggest an alternative: Let the fixed effects capture any unobserved heterogeneity that is outside the firms' managements control. Thus, we speak of any heterogeneity that affects the overall level of costs but not the firms' relative efficiency scores. Examples might be: Country effects, specialization effects, ownership effects or time effects. Actually, with our data at hand, it is obviously recommendable to estimate vertical time-specific location shifts of the conditional Quantile Regression lines. So we estimate α_t for all $t = 1, ..., T$:

$$\min_{(\beta,\alpha)\in\mathcal{R}^{(K+1)+T}} \sum_{q=1}^{Q}\sum_{i=1}^{N}\sum_{t=1}^{T} w_q \rho_{\tau_q}(c_{it} - \alpha_t - x'_{it}\beta(\tau_q)) \tag{6.9}$$

This is equivalent to the assumption of a time-invariant shape (but a time-variant location) of the conditional distribution $c_{it}|x_{it}, \alpha_t$. The respective quantile slope and intercept parameters reflect the technical aspects of highly efficient versus less efficient production and remain the same in all years. The comparison of high quantile cost functions with low quantile cost functions gives an impression of how inputs and outputs should be optimally allocated in the firms' production processes. We only allow the level of costs, i.e. the vertical shift α_t, to vary from year to year. So when measuring inefficiency as vertical distance from a best-practice low quantile cost function (Behr, 2010), shifted every year, we control for unobserved environmental factors that affect all firms' costs in the same way in the particular year. This is the additional benefit we earn from the use of longitudinal data in Quantile Regression efficiency estimation.

As we have T observerations per firm, measuring the distance between a low quantile cost frontier in t and the observations delivers T unrestricted inefficiency scores u_{it} for every i. Now the question arises how we can consolidate the information contained in multiple inefficiency scores to gain easily accessible results.

- On the one hand, it is possible to estimate an efficiency trend function $u_{it} = f(t, t^2, ...)$. According to the panel length, even firm-specific trends f_i are feasible. In this case, the empirical distribution of the trend-parameters might be of interest.

- On the other hand, one can refer to the basic idea of the Distribution Free Approach DFA (Berger, 1993): Assuming time-invariant inefficiency for every i, the mean inefficiency scores $1/T \sum_t u_{it}$ might be implicitely corrected for temporarily positive or negative influences of the random noise. The author argues on the grounds of the supposition that good and bad luck balance against each other in retrospective after a few years.

As we see, multiple possibilities exist. To put forward a robust estimation of time-variant efficiency, using the full information obtained on $c_{it}|x_{it}$, we additionally propose the following 'efficiency sector concept': As Koenker (2004) suggests estimation of $Q = 3$ Quantile Regression lines simultaneously, the conditional distribution $c_{it}|x_{it}, \alpha_t$ splits into four stacked sectors (that are vertically shifted every year). For example, the most cost efficient sector lies below the $\tau = 0.10$ regression line, the fairly efficient sector between $\tau = 0.10$ and $\tau = 0.5$, the rather inefficient sector above the median and below the $\tau = 0.90$ regression line, as well as the clear-cut inefficient sector above $\tau = 0.90$.

Surely the precision of the sector concept can be improved by estimation of more regression lines. Howsoever, the information contained in the sector location of every residual is more detailed than a single efficiency trend, assuming an identical behaviour for all firms in the dataset; and less detailed than firm-specific trends that might provide too much unstructured information.

In figure 6.1 we graphically illustrated the basic idea by means of a stylized two-dimensional problem. If we let $N = 10$ and $T = 2$, the observations in $t = 1$ are marked as triangles, in $t = 2$ as squares. The slope as well as the (relative) intercept coefficients equal in both years. Remember that in cross-sectional Quantile Regression about $\tau \cdot 100\%$ of the observations lie below the $\tau-$ regression line[8]. Now note that the basic property of panel Quantile Regression does not require $\tau \cdot 100\%$ of the observations to lie below the regression line in every single year, but jointly over all years. This is important as it allows us in our sector concept to identify more efficient and less efficient years.

[8]Due to the employed optimization procedure, at least K observations lie exactly *on* the regression line.

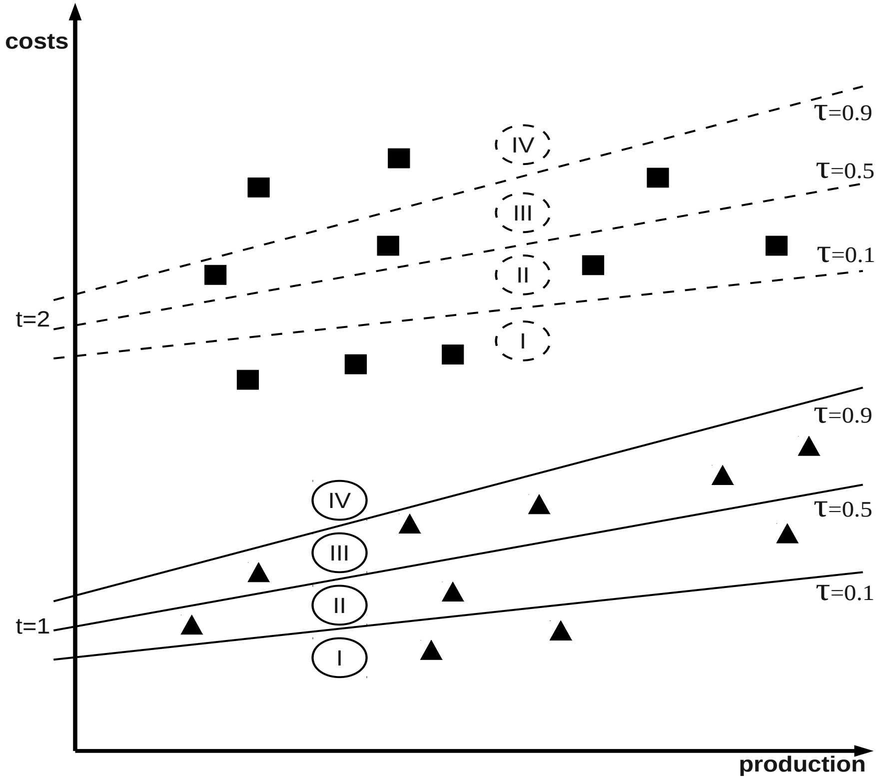

Figure 6.1: Efficiency sector concept – schematic

The design matrix

We would like to complete this section with some technical details concerning the optimization procedures. In our example, and with $Q = 3$, the design matrix D of the problem (6.9), with $\left(x_{it}\right)$ the matrix of independent variables including intercept, dimension $(NT \times (K+1))$, and I a vector of ones of length N indicating the dimension of the fixed year effects, takes the form (Koenker, 2004):

$$D = \left(\begin{array}{ccc|cccc}
 & & & w_1 \cdot I & 0 & \cdots & 0 \\
w_1 \cdot \left(x_{it}\right) & \left(0\right) & \left(0\right) & 0 & w_1 \cdot I & & \vdots \\
 & & & \vdots & & \ddots & 0 \\
 & & & 0 & \cdots & 0 & w_1 \cdot I \\ \hline
 & & & w_2 \cdot I & 0 & \cdots & 0 \\
\left(0\right) & w_2 \cdot \left(x_{it}\right) & \left(0\right) & 0 & w_2 \cdot I & & \vdots \\
 & & & \vdots & & \ddots & 0 \\
 & & & 0 & \cdots & 0 & w_2 \cdot I \\ \hline
 & & & w_3 \cdot I & 0 & \cdots & 0 \\
\left(0\right) & \left(0\right) & w_3 \cdot \left(x_{it}\right) & 0 & w_3 \cdot I & & \vdots \\
 & & & \vdots & & \ddots & 0 \\
 & & & 0 & \cdots & 0 & w_3 \cdot I
\end{array}\right) \tag{6.10}$$

The corresponding vector of the dependent variable c^*_{it} has the length $Q \cdot NT$:

$$c^*_{it} = \begin{pmatrix} w_1 \cdot \left(c_{it}\right) \\ w_2 \cdot \left(c_{it}\right) \\ w_3 \cdot \left(c_{it}\right) \end{pmatrix} \tag{6.11}$$

Note that an OLS conditional mean regression of c^* on D cannot be solved by $(D'D)^{-1}(D'c^*)$ as the product matrix $(D'D)$ is singular due to the existence of multiple intercept indicators. Users of the R language may simply pass c_{it}, x_{it} and a vector containing fixed effects indices to the `rq.fit.panel()`-function available on the homepage of Roger Koenker. Inference basically is not available and can at best be derived from bootstrapping procedures (Kato et al., 2010).

6.3 Empirical Results

Our exemplifying application covers the field of bank efficiency. It is a common procedure to estimate cost functions for banking firms. Concerning the inputs and outputs of a production process in banking, we would like to abstain from the controversial discussion at this point. Instead, we stick to one of the most recent studies, namely Girardone et al. (2009), and model the production process according to the Intermediation Approach (Sealey and Lindley, 1977).

6.3.1 Data preparation

We accessed a Bankscope[9] balance sheet and income statement dataset of commercial banks in two countries, i.e. Germany (DE) and the USA (US). We prepared all variables necessary to estimate usual cost functions in banking, listed in the so-called 'Bankscope Global Format'. When we restrict our observations to hold a positive figure > 100T USD in total assets and a plausible figure in all other variables we are interested in, we count a total of 3774 observations. The numbers sum up over the years 2000 – 2007 as given in table 6.1.

	2000	2001	2002	2003	2004	2005	2006	2007	SUM
DE	146	147	143	140	141	141	143	129	1130
US	388	377	352	340	314	312	296	265	2644
SUM	534	524	495	480	455	453	439	394	3774

Table 6.1: Numbers of observations

The application of methods for panel data usually makes high demands on the number of years each bank is observed (individual panel length). Table 6.2 shows how many of the banks are observed in how many years.

Fortunately, most of the banks are observed over the full panel length. So just as we simplified our notation to the case of a balanced sample, we now restrict our analysis to 323 banks, 87 in Germany and 236 in the USA.

[9]Bureau van Dijk, www.bvdep.com.

	1	2	3	4	5	6	7	8	SUM
DE	7	18	10	13	10	14	25	87	184
US	24	32	29	30	18	28	29	236	426
SUM	31	50	39	43	28	42	54	323	610

Table 6.2: Number of banks observed over a period of 1-8 years

All data are given in thousands of US-Dollar; EUR values were converted by means of a single exchange rate to exclude floating effects. Moreover, to avoid the measurement of inflation effects over the years, we corrected all nominal values by OECD consumer price indices with the base year 2005.

To ensure basic data consistency, annual growth rates of total assets were calculated. We detected some single outliers (implausible annual growth rates of about 6000%) and excluded the respective banks from the sample. The distribution of the growth rates of the remaining banks are plotted in figure 6.2. The box represents the interquartile-range, the horizontal dash being the median.

We see that a predominant share of banks exhibit a positive growth rate. In the USA, not even the first quartile of banks fall below the 0% threshold. In Germany, especially in 2002 and 2003, nearly 50% of the banks shrank in terms of total assets.

In figure 6.3, we prepared an overview illustrating the distribution of bank sizes in Germany and the USA, each in 2000 and 2007. On the left-hand side, the respective densities of the distributions of log total assets show that banks in the USA tend to be bigger than banks in Germany. Some huge banks do exist, though, in Germany (e.g. Deutsche Bank, Commerzbank, Dresdner Bank). So obviously, the distribution of total assets even in logs is still right-skewed and not normal.

Not surprisingly, the existence of very big banks in Germany is reflected in common inequality measures like the Lorenz curve and the related Gini coefficient (right-hand side of figure 6.3): The inequality of total assets (not in logs) per bank in Germany is greater than in the USA, but slightly decreasing from 2000 to 2007. Contrary, inequality in the USA tends to increase. Together with our findings in figure 6.2, this

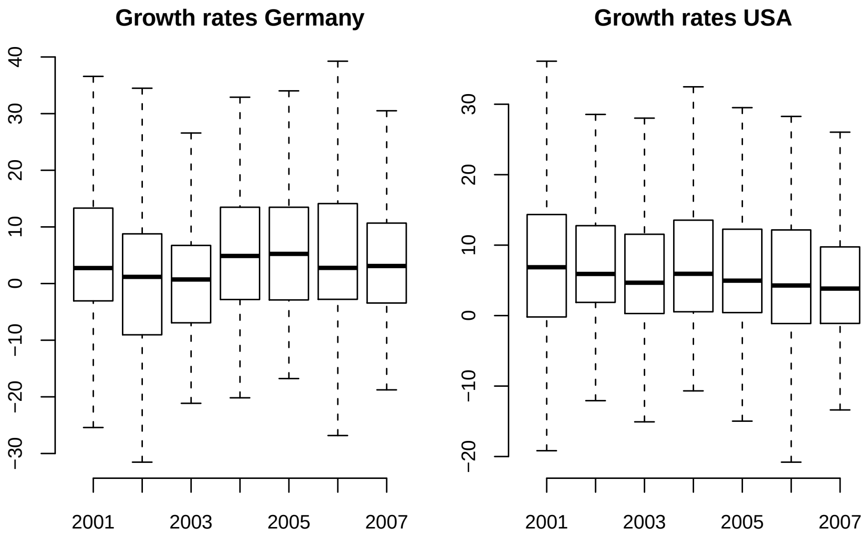

Figure 6.2: Box plot of annual growth rates

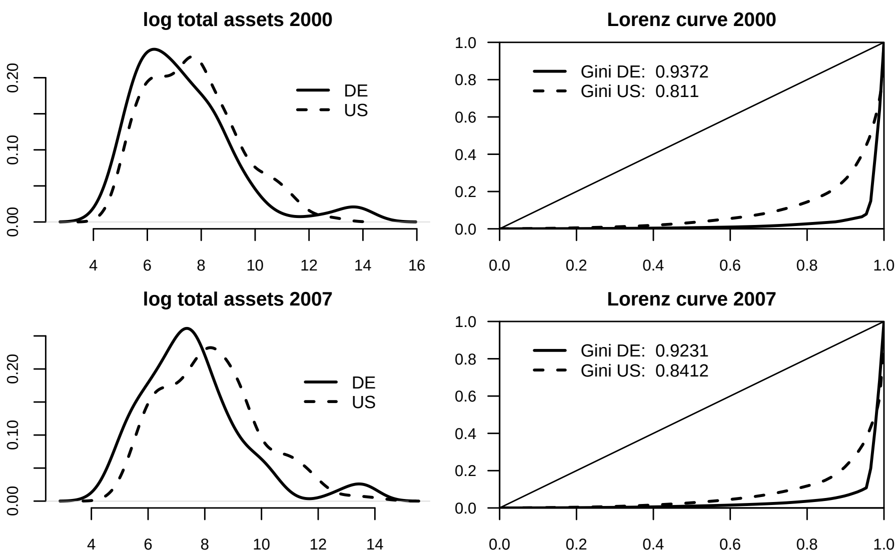

Figure 6.3: Density of log total assets and the respective Lorenz curves

gives us a hint that big banks in the USA are growing faster than small- and medium-sized banks, whereas in Germany, big banks are likely to grow more slowly than smaller banks.

6.3.2 Cost functions

According to the Intermediation Approach (Sealey and Lindley, 1977) and in line with recent literature (Girardone et al., 2009), the variables of an indirect cost function (Coelli et al., 2005, chapter 2.4) are defined as follows:

- c total costs: Interest expenses + other operating expenses + personnel expenses
- y_1 output 1: Loans
- y_2 output 2: Other earning assets
- w_1 price of labour: $\frac{\text{personnel expenses}}{\text{total assets}}$
- w_2 price of fixed assets: $\frac{\text{other operating exp. (incl. depreciations)}}{\text{total fixed assets}}$
- w_3 price of deposits: $\frac{\text{interest expenses}}{\text{total deposits}}$

Descriptive statistics on the mean, median and standard deviation for either country separately are given in table 6.3.

	c	y1	y2	w1	w2	w3
mean DE	1170.19	14978.56	17009.56	0.03	0.81	0.06
median DE	65.91	614.66	340.54	0.02	0.68	0.05
sd DE	5443.43	70462.08	76565.41	0.03	1.03	0.04
mean US	916.60	9412.93	5247.29	0.02	3.38	0.04
median US	137.97	1557.54	582.08	0.02	1.07	0.04
sd US	3528.78	32483.25	32220.64	0.02	15.61	0.06

Table 6.3: Descriptive statistics, all variables

We see that the ratio of mean to median is very large. This is a typical characteristic of variables describing banks' key figures. Some authors mistake extremely large values for outliers, and therefore cut large banks out of the dataset. We prepared another graphic which contains boxplots illustrating the distribution of every variable over time (figure 6.6 in the appendix).

We see that both outputs y_1, y_2 tend to increase over the years. Note that we already corrected all values for inflation, so that we have to do with real growth. While the price of labour and the price of fixed assets slightly decrease over the years, the median interest paid on deposits is U-shaped, just like total costs K, and reflects the interest levels (e.g. EURIBOR) in the respective years.

Due to an insufficient number of observations and to ease the interpretation of estimated parameters we abstained from estimation of a more flexible translog cost function. But even a Cobb-Douglas cost function must not be estimated without certain restrictions. Actually, regularity conditions require that the input price elasticities sum up to 1. So we divided all prices and costs by the price of total deposits, w_3 (Lang and Welzel, 1996):

$$\begin{aligned}
c =& e^{\beta_0} \cdot y_1^{\beta_1} \cdot y_2^{\beta_2} \cdot w_1^{\beta_3} \cdot w_2^{\beta_4} \cdot w_3^{\beta_5} \\
&\text{s.t.} \quad \beta_3 + \beta_4 + \beta_5 = 1 \\
&\Leftrightarrow \quad \beta_5 = 1 - \beta_4 - \beta_3 \\
c =& e^{\beta_0} \cdot y_1^{\beta_1} \cdot y_2^{\beta_2} \cdot w_1^{\beta_3} \cdot w_2^{\beta_4} \cdot w_3^{(1-\beta_4-\beta_3)} \\
\left(\frac{c}{w_3}\right) =& e^{\beta_0} \cdot y_1^{\beta_1} \cdot y_2^{\beta_2} \cdot \left(\frac{w_1}{w_3}\right)^{\beta_3} \cdot \left(\frac{w_2}{w_3}\right)^{\beta_4} \qquad (6.12)
\end{aligned}$$

To detect 'real outliers' in the context of a regression model, we performed regression deletion diagnostics[10]. Estimating the pooled linear regression with and without every single observation, this (Jackknife-)procedure reveals the observations which strongly bias each of the estimated parameters. We plotted the corresponding boxplots indicating the respective bias in absolute values on the ordinate axis to graphically identify the outliers (figure 6.7 in the Appendix). Obviously, we find a

[10] So-called 'leave-one-out deletion' (Cook and Weisberg, 1982). The corresponding procedures are available within the R Environment.

considerable influence on the intercept parameters. To tell the truth, this comes in rather handy, because Quantile Regression with fixed effects is a tool to disentangle multiple forms of heterogeneity on the y-axis (economic and econometric heterogeneity, compare our discussion in section 6.2).

Finally, we splitted the dataset into two subsets: One for Germany (DE) and one for the USA (US). We are well aware that under the assumption of country-specific cost frontiers banks do not compete against the same benchmark, and consequently efficiency comparisons between both countries are not justified. On the other hand, assuming identical benchmark technologies across numerous countries requires an appropriate set of environmental control variables (Berger, 2007). To circumvent the implications of this discussion in our study primarily focused on methodology, we use the opportunity to apply the method to two different datasets. Comparing the results will not reveal any information about efficiency advantages of banking firms in one country over the other country, but it will provide more insight into the economic particularities of the respective country.

6.3.3 Estimation

As we are about to estimate Quantile Regressions with fixed effects, we first estimated the associated conditional mean models $E(c_{it}|x_{it}, \alpha_t)$ to obtain a first impression of the expected parameters. Tables 6.4 and 6.5 show the slope parameters for the German and the US dataset, respectively. A direct comparison of both countries reveals that the production of one percent more loans (y_1) raises costs in Germany by 0.63%, but in the USA by only 0.55%. Regarding other earning assets (y_2), costs in Germany increase less than in the USA. The estimated coefficients for w_1 and w_2 are adjusted to meet the regularity conditions given in equation (6.12). This means that the missing parameters w_3 can be calculated as $1 - w_1 - w_2$. As expected, all cost elasticities are positive, telling us that the basic assumptions concerning the elasticity of substitution are in conformity with the data[11].

[11]Note that the restriction in equation (6.12) basically does not prohibit estimation of negative price elasticities. In Behr (2010, Figure 9), we see that the same

	Estimate	Std. Error	t-value	Pr(>|t|)
y1	0.63	0.01	63.05	0.00
y2	0.35	0.01	37.13	0.00
w1r	0.21	0.03	7.54	0.00
w2r	0.12	0.02	5.87	0.00

Table 6.4: Fixed effects estimation DE

	Estimate	Std. Error	t-value	Pr(>|t|)
y1	0.55	0.01	76.37	0.00
y2	0.42	0.01	60.04	0.00
w1r	0.27	0.01	22.64	0.00
w2r	0.45	0.01	47.37	0.00

Table 6.5: Fixed effects estimation US

The corresponding time-specific intercept parameters are given in table 6.6. They represent the location shift of the conditional mean functions. So we see that the cost level is U-shaped in the USA, with costs in the year 2004 being lower than in all other years. In Germay, the cost level is lowest in 2000, and tends to increase in the subsequent years. The maximum is in 2007.

The model in (6.9) was estimated by means of Koenker's prototype function. $Q = 3$ quantiles $\tau = (0.1, 0.5, 0.9)$ with Tukey's trimean weights $w = 0.5 - |\tau - 0.5|$ were simultaneously estimated. The results are given in table 6.7. The first six parameters are indexed by $\tau_1 = 0.1$, the next six parameters by $\tau_2 = 0.5$ and the last six by $\tau_3 = 0.9$. Below, the year-specific location shifts are given.

We see that the parameters slightly differ from quantile to quantile, but the relative influence of the explanatory variables remains stable. So in Germany, in line with the conditional mean regression above, we have

parameters at extremely high conditional quantiles (characterizing inefficient firms) turn out to be negative. In this case, the data contradict the underlying assumption of a substitutional cost function that is nondecreasing in input prices.

	DE	US
2000	1.1672	0.2779
2001	1.1889	0.2370
2002	1.1828	0.1477
2003	1.1737	0.1187
2004	1.1774	0.1032
2005	1.1728	0.1547
2006	1.1922	0.2412
2007	1.1930	0.2667

Table 6.6: Fixed effects parameters

$\beta_{y_1} > \beta_{y_2}$ and $\beta_{w_3} > \beta_{w_1} > \beta_{w_2}$ for all three quantiles. In comparison with table 6.4, the conditional mean parameters rather resemble the lower quantiles. As β_{w_3} is rather large, banks in Germany seem to suffer most from increasing deposit rates. Refering to figure 6.6, this proposition is corrobarated by the syncronous development of deposit rates w_3 and costs c.

In the USA, we have also $\beta_{y_1} > \beta_{y_2}$ for all quantiles and for the expectation (table 6.5). But the relative influence of the price elasticities differs among the quantiles. What we can state is that the deposit rate does not exert much influence on costs as it does in Germany. Moreover, especially in the median regression as well as in the conditional mean regression, all cost elasticities are rather balanced.

From cost theory, we know that the sum of the parameters β_{y_1}, β_{y_2} reveals some information about the underlying dual production function. So if $\beta_{y_1} + \beta_{y_2} > 1$ or < 1 in a cost function[12], the corresponding production process exhibits decreasing/increasing returns to scale, which might be seen as an indication that firms are too large/small (Coelli et al., 2005, p.18 f.). Obviously, in our case, $\beta_{y_1} + \beta_{y_2} \approx 1$, so banks in both countries seem to exhibit constant returns to scale.

The reported year effects at the end of table 6.7 differ from the effects

[12]If standard errors of the parameters were available, tests of $\beta^{y_1} + \beta^{y_2} = 1$ would be possible.

	$\tau_1 = 0.1$		$\tau_2 = 0.5$		$\tau_3 = 0.9$	
	DE	US	DE	US	DE	US
Int	0.62	0.42	1.33	0.40	2.47	0.72
y1	0.66	0.76	0.64	0.73	0.57	0.50
y2	0.38	0.31	0.34	0.32	0.35	0.43
w1	0.24	0.41	0.17	0.37	0.28	0.15
w2	0.11	0.21	0.10	0.30	0.09	0.58
w3	0.65	0.37	0.74	0.32	0.64	0.27
	FE	2000	-0.24	-0.29		
	FE	2001	-0.22	-0.33		
	FE	2002	-0.21	-0.39		
	FE	2003	-0.19	-0.41		
	FE	2004	-0.18	-0.44		
	FE	2005	-0.18	-0.40		
	FE	2006	-0.18	-0.34		
	FE	2007	-0.19	-0.31		

Table 6.7: Regression results DE and US

in tables 6.4 and 6.5 in terms of absolute values. But a closer look reveals that all values are more or less just shifted relative to the conditional mean FE parameters. So as a result, we again observe that the cost level in Germany tends to increase, while the cost level in the USA is U-shaped. Remember that U-shaped total costs in Germany are captured by the high influence of w_3. So it is a little bit surprising that we see the particular development of costs being explained by β_0 in the one case and by α_t in the other case.

6.3.4 Efficiency

The 'traditional' approach to inefficiency in QR is to calculate the residuals u_{it} from the lower 0.1 cost frontier. As all variables are in logs, $u_{it} = \log(c_{it}^{actual}|x_{it}, \alpha_t) - \log(c_{it}^{frontier}|x_{it}, \alpha_t)$ is inefficiency measured as percentaged cost surplus. We plotted all u_{it} per year in figure 6.4.

10% of all points are negative, because 10% of the observations undercut the cost target. The solid lines for both countries indicate the location of fitted linear time trend functions. In Germany, no positive or negative efficiency trend exists, whereas in the USA, overall inefficiency tends to increase. But as expected, the evidence found in this figure is rather sparse: Does inefficiency for all firms in Germany remain constant, or do ranking positions change? Is it due to single firms in the USA that overall inefficiency increases? To gain some deeper insight, we additionally estimated firm-specific linear trends of the form $u_{it} = \gamma_{0i} + \gamma_{1i} \cdot year_t + noise_{it}$ for all i. The distributions (histogram and kernel density estimation) of the N trend parameters $\hat{\gamma}_1$ for both countries are plotted in figure 6.5.

As the vertical dashed lines represent the interquartile-range, we see that trend parameters for banks in both countries spread equally, whereas in the USA, some 'outliers' with strongly increasing inefficiency exist. Consequently, it seems to be due to single banks in the USA that overall inefficiency in figure 6.4 increases, too. But for the entirety of banks, we can assume time-invariant inefficiency in both countries on average.

In section 6.2.2, we described our concept of efficiency sectors. As a consequence of the existence of $Q = 3$ regression lines, we obtain $NT \cdot 3$ residuals in the course of the estimation procedure. In R-code, a function to assign the respective efficiency sector out of the three deviations per observation from the τ_1, τ_2 and τ_3 quantile function may look like:

```
sector <- function(x) { # input x: 3 deviations tau1, tau2, tau3
          result <- NA
          if(x[1]<=0 & x[2]<0 & x[3]<0) {result <- 1}
          if(x[1]>0 & x[2]<=0 & x[3]<0) {result <- 2}
          if(x[1]>0 & x[2]>0 & x[3]<=0) {result <- 3}
          if(x[1]>0 & x[2]>0 & x[3]>0) {result <- 4}
          return(result)
}
```

The resulting numbers of observations per sector and year are given in tables 6.8 and 6.9. In Germany, we can now see in more detail that the number of highly efficient banks increases, while the number of less efficient banks (sector 2) decreases. In sectors 3 and 4, the number of observations remains relatively stable. On the other hand, in the USA,

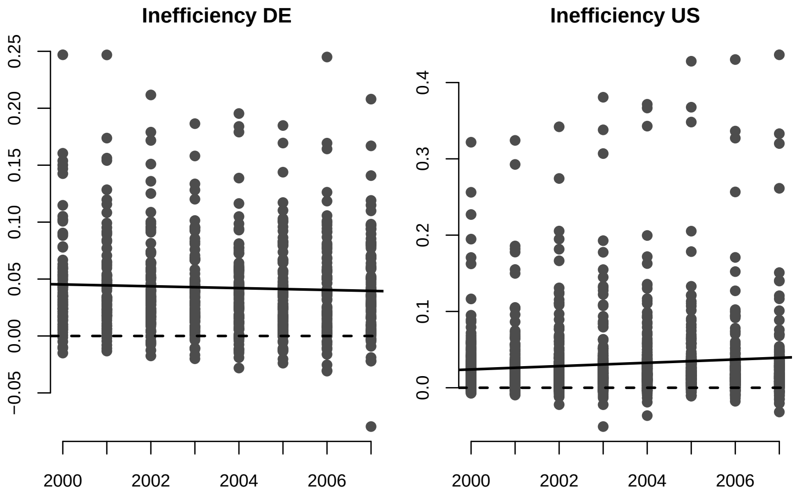

Figure 6.4: Inefficiency, deviations from $\tau = 0.1$ frontier

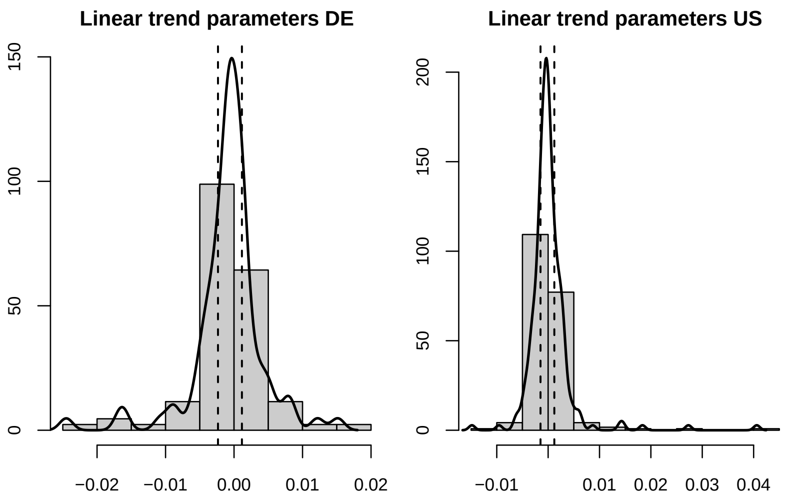

Figure 6.5: Distribution of N inefficiency trend parameters (DE, US)

we get the impression that the medium sectors 3 and 4 become less important in favour of the extreme sectors 1 and 4.

Unfortunately, on the basis of these tables, we cannot draw any conclusions about the interchange of banks in the respective sectors. In fact, all sectors could be assigned every year to different banks, as this is just a static survey. To capture the dynamic interchange, we prepared another table (6.10 in the appendix). It is an example of an extended analysis made possible by application of the sector approach: The columns represent the four sectors in the year 2000. In the rows, we find the number of banks having changed from the respective sector in 2000 to another (or the same) sector in the subsequent years. So for example, we learned from table 6.8 that we have five banks in sector 1 in Germany in 2000. Table 6.10 now tells us that two of these banks fell off in efficiency (sector 2) in 2001 and three banks remain in sector 1. Until 2007, another bank further declines to sector 3. As a matter of course, the table can be extended to account for every year as a reference.

	2000	2001	2002	2003	2004	2005	2006	2007	sum	percent
sector 1	5	7	8	7	12	10	11	10	70	10.06
sector 2	39	37	36	36	30	33	32	34	277	39.80
sector 3	32	34	34	38	37	36	36	32	279	40.09
sector 4	11	9	9	6	8	8	8	11	70	10.06

Table 6.8: Classification DE

	2000	2001	2002	2003	2004	2005	2006	2007	sum	percent
sector 1	22	20	20	26	23	19	25	32	187	9.91
sector 2	95	98	97	91	94	99	93	87	754	39.96
sector 3	100	98	102	97	100	97	86	78	758	40.17
sector 4	19	20	17	22	19	21	31	39	188	9.96

Table 6.9: Classification US

6.4 Conclusion

In this study we discussed the Quantile Regression approach to inefficiency measurement. Although cross-sectional methods have recently been established in literature, repeated observations have remained unused so far. We picked up the panel data methods suggested by Koenker (2004), and discussed the new opportunities in the course of efficiency estimation. In particular, we had to cope with two problems: First, the methodology of Quantile Regression with longitudinal data is technically not mature. This is closely related to a fundamental difficulty when interpreting the parameters, especially when disentangling multiple intercept parameters. The second problem was the transfer of the model to the field of efficiency measurement. Nevertheless, we clarified the particular advantages of the new approach over the customary components: (i) conditional mean regression with firm-specific fixed effects indicating relative inefficiency and (ii) robust QR cost frontiers without the benefit of repeated observations. Using multiple Quantile Regression lines simultaneously provides the basis for a robust efficiency sector concept, using the full information we estimated to characterize the conditional distribution of costs. In the course of our exemplifying application, we estimated cost efficiency for German and US commercial banks during the period 2000 till 2007. We demonstrated that the efficiency sector concept does in fact work and, moreover, opens up new possibilities to asses (bank) efficiency.

6.5 Appendix

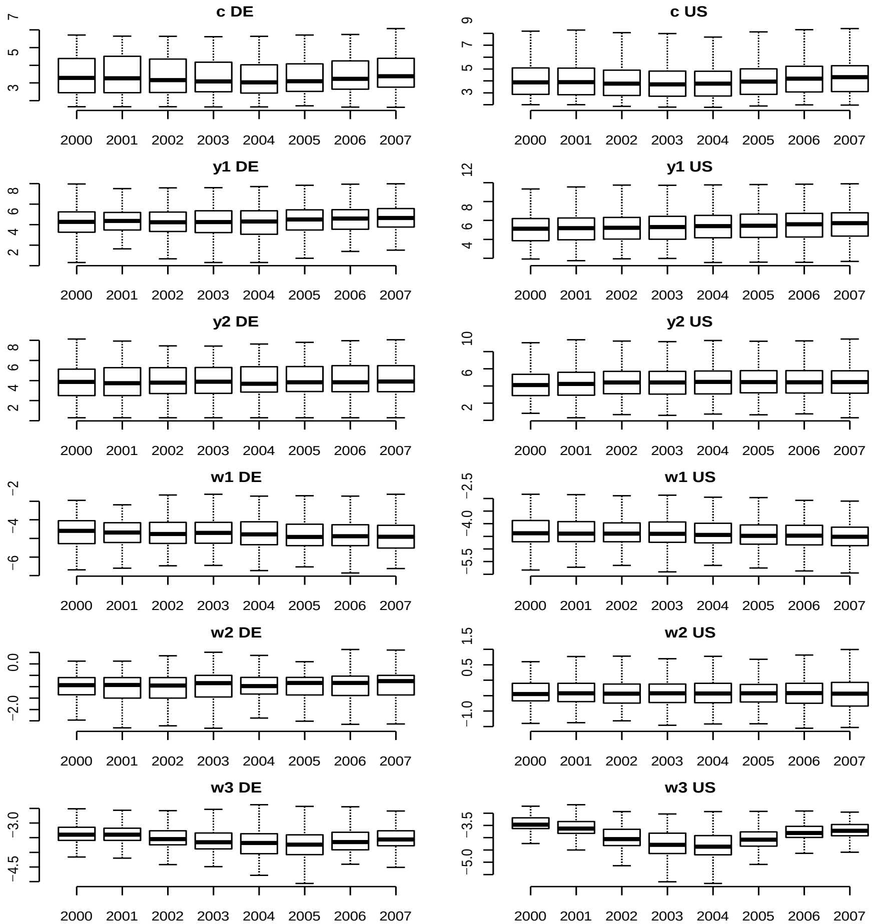

Figure 6.6: Boxplots, all values in logs

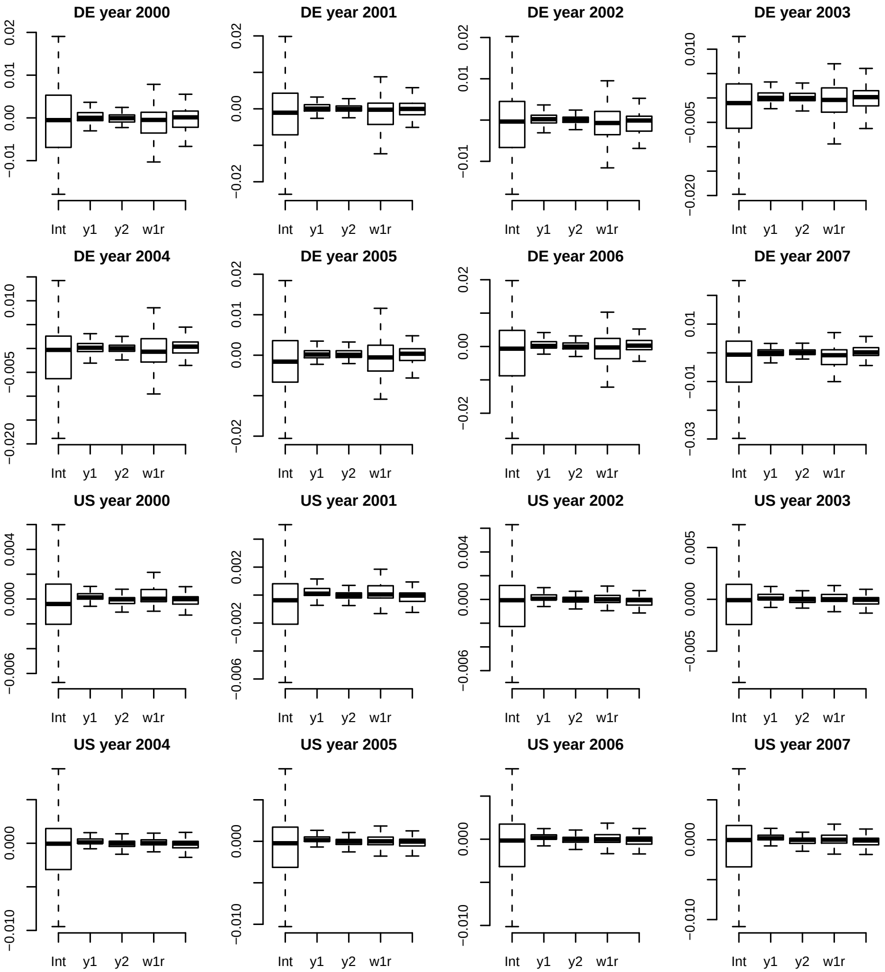

Figure 6.7: Regression deletion diagnostics

	00;1	00;2	00;3	00;4	00;1	00;2	00;3	00;4
01;1	3	3		1	10	8	2	
01;2	2	30	5		11	68	18	1
01;3		5	27	2	1	17	75	5
01;4		1		8		2	5	13
02;1	2	5		1	8	9	3	
02;2	2	28	6		12	61	23	1
02;3	1	5	25	3	2	24	68	8
02;4		1	1	7		1	6	10
03;1	2	4		1	8	13	5	
03;2	2	25	8	1	9	52	29	1
03;3	1	10	22	5	4	28	57	8
03;4			2	4	1	2	9	10
04;1	2	8	1	1	8	11	4	
04;2	2	21	6	1	8	57	27	2
04;3	1	10	22	4	6	26	60	8
04;4			3	5		1	9	9
05;1	2	7		1	7	10	2	
05;2	2	23	6	2	10	57	29	3
05;3	1	8	24	3	5	26	59	7
05;4		1	2	5		2	10	9
06;1	3	7	1		9	14	2	
06;2	1	20	9	2	7	53	30	3
06;3	1	11	20	4	5	23	51	7
06;4		1	2	5	1	5	16	9
07;1	3	5	2		12	16	4	
07;2	1	23	8	2	4	49	30	4
07;3	1	11	18	2	5	19	47	7
07;4			4	7	1	11	19	8

Table 6.10: Contingency tables DE (left) and US (right)

7

Chapter 7

Concluding remarks

7.1 The scope of this survey

The amount of literature on bank efficiency accrues from the need to objectively assess the performance of banks against the background of a rapidly changing institutional environment. Especially changes in the legal framework (e.g. the introduction of EU law in former socialist countries), the privatization of banks in a market economy, the socialization of banks in the light of systemic risks, the loss of competition resulting from merger activities, etc. give authors occasion to estimate the respective impact on the efficiency of banks.

So as a result, the literature on bank efficiency estimation is one of the most vividly growing branches of economic science. Readers trying to get acquainted with this field of research rely on the existence of comprehensive surveys which encompass the multiplicity of economic issues, models, and methods covered. We found such pioneering works in Berger and Humphrey (1992), Berger et al. (1993), Berger and Humphrey (1997), Berger and Mester (1997), Berger (2007).

In order to characterize our own contributions to this literature, we subdivided the existing studies into two lines for the sake of convenience: First, studies dealing with several kinds of real economic problems mentioned above are issued, for instance, by authors from banking supervision authorities (central banks), the International Monetary Fund, the OECD, or the Bank for International Settlement. The efficiency methods typically applied in the course of these studies are standard procedures implemented in econometric software packages. On the other hand, the second type of studies focuses on innovations in exactly these methods. The authors are – in the broadest sense – econometricians who need an exemplifying application to illustrate the advantages of their proposals. We did not find any explicit hints at macroeconomic problems that are being addressed in the course of these studies, but the announcement that the new models are applied to bank efficiency data (Kumbhakar and Tsionas, 2008, Greene, 2002, 2005b). So we can state that the scope of bank efficiency literature goes even beyond questions regarding only the economic performance of banks but is also a reference object for methodical innovations.

In this survey we mainly contributed to the latter field of literature, i.e. the econometrics of bank efficiency estimation. Nevertheless, we first took into account the fact that econometric models are always part of a scientific procedure called the 'Scheme of Adequation' (figure 3.1 in section 3.2.1). This scheme put us in a position to substantiate the existence of an 'information deficit observed' (Brachinger, 2007) between the general operationalization procedure and the interpretation of the results in empirical studies mentioned in the former case. In the following we focused on methodical innovations, i.e. the link between a formal problem and a formal solution, without expatiating on the interpretation of the formal solution against the background of a real economic problem. Moreover, the operationalization of a commonly used production model in banking (Sealey and Lindley, 1977) adopted from the literature was employed to ensure the basic comparability of our formal results with preceding studies.

7.2 Problems and solutions

7.2.1 Adequation

In the course of our methodical modules we used banking data to illustrate our proposals. It was our aim to stick as closely as possible to the existing operationalization procedures in the literature. Using Bankscope data with unified identifiers, as most of the other authors do, provides a very convenient way to reproduce the classic production models. Nevertheless, we could not do away with the question in how far the estimated results contribute to the explanation of macroeconomic problems in the banking sector. In particular, we read about the fundamental 'shotgun marriage' reproach on the part of Stigler (1976). Although this is a general criticism regarding every form of efficiency measurement, we had to ask ourselves whether it holds for the field of bank efficiency estimation, too.

Stigler refers to situations in which the institution being assessed is not aware of the performance standards it is to achieve or does not agree to that benchmark. In the case of banks we learned that the usual standard that had been used for years was the transformation of deposits to loans and investments at minimum costs/maximum profit. Whatever economic issue was addressed in the course of a study, the given standard was applied.

Obviously, literature has come to the conclusion that banks cannot agree to the production model according to Sealey and Lindley (1977) any more. In recent decades, the product range of banks has considerably diversified. As a result, banks, for instance, nowadays engage in risk management activities and the trade with derivative products. Most of these activities are not captured by balance sheet items, but are recorded 'off-balance-sheet' (OBS). Authors make considerable endeavours to adapt the classic operationalization by means of new data in order to meet the banks' own notion of an appropriate standard.

We considered this approach to be problematic. To substantiate our supposition, we had recourse to the Scheme of Adequation (Grohmann, 1988): According to this fundamental scheme every empirical analysis starts with the formulation of the economic problem. We learned that the process of the formulation is determined by concepts that have to be understood as ideal types. In other words: Economic problems are not 'observable' in the sense of 'countable'. As a matter of fact, it falls to the researcher to achieve a consensus in the way he describes the problem. As this is typically a challenge which cannot be undervalued Weber (1988) lays down the 'rules' which form the requirements for a consensus among the readers. In particular, we identified four requirements for every ideal-type formulation (section 3.2.2):

1. Abstraction of reality (Utopia)
2. Absence of personal judgements
3. Imagination of typical conditions
4. 'Means to an end'

So in the case of banks, we first had to ask ourselves what the basic 'ideal-type' characteristics are like. Or, more specifically, what are banks supposed to perform? Only against this standard can we start an empirical investigation on the economic problem why banks do *not* perform as expected. We raised some suggestions about the expectations one might impose on banks and considered how these aspects might be quantified in 'concrete reality' (section 2.2). We learned that the standards associated with banks depend on the respective interest group and are by no means congruent among the groups. This makes it all the more important for authors to precisely lay down the basic goals they pursue.

Amazingly, we never read about similar considerations in bank efficiency literature. So consequently it is advisable to pose the question whether the operationalization, i.e. the image of reality we encounter in literature reflects the ideal-type conceptualization of banks immanent in the formulation of real economic problems. In fact, to our knowledge, the model by Sealey and Lindley (1977) is the last explicit ideal-type formulation of bank activities we came across in the field of bank efficiency literature. We already mentioned that this model is not considered to be appropriate any more, but simple variations in the operationalization do not constitute a new ideal type, because they do not meet the 'objectivity' requirements mentioned by Max Weber. Moreover, we read in studies (Tortosa-Ausina, 2002) that in fact data availabilty determines our model of banks and not the consensus among the readers/researchers. Concretely: We observed bank efficiency studies starting with the operationalization process and not with the ideal-type description of bank activities constituting the economic problem.

So the results of our study definitely lead to the conclusion that the principles of Adequation are violated. Consequently, the interpretation of the formal results against the background of a real economic problem cannot be assessed by the scientific standards postulated in Weber (1988). It is exactly this 'adequation gap' which can be perceived as an 'information deficit observed' (Brachinger, 2007) by the readers. The solution to this problem is self-evident: A study assessing the performance of banks has to start with a description of the basic tasks banks are supposed to fulfill – and this does not necessarily mean that any economic problem always has to do with the transformation process of deposits to loans at minimum costs. In our opinion, this is the natural option to put the reader in a position where he can first judge the consensus on the description of the problem, and second where he can evaluate the adequacy of the operationalization.

7.2.2 Methods

The first methodical issue we addressed was the question whether a Stochastic Frontier model should be estimated by means of a numerical Maximum Likelihood procedure or rather by a two-step Method of Moments approach. In the context of cross-sectional data we performed extensive simulation studies under various distributional assumptions, including misspecification scenarios, thereby contributing to the existing literature (Olson et al., 1980, Coelli, 1995). On the basis of the respective results we derived the recommendation to favour the MOM approach in small samples with little inefficiency. A parametrical decision support was given in the form of 'rules of thumb'.

In the next step of our analysis we turned to the treatment of panel data. In order to use the longitudinal structure of the data the modification of a random effects approach by Kumbhakar (1990) allowed us to estimate group-specific inefficiency trend parameters. This proved to be an obvious compromise between economic plausibility and technical feasibility: The original model comes from the assumption of a single common trend parameter which is surely easy to estimate but hardly reveals additional economic information. The other extreme, i.e. firm-specific trend parameters (Cuesta, 2000), probably demands too much information, namely in practice the estimation of several hundred parameters in the course of a numerical optimization procedure.

Furthermore, the panel data environment opens the possibility to overcome the a priori-specification of a distributional assumption concerning inefficiency. We learned that inefficiency is basically unobservable so that there does not exist any reason to prefer a specific distribution to another distribution. And as in fact the distributional assumption directly determines the estimated efficiency scores, the results seem to be arbitrary. But Greene (1990) showed that the relative efficiency scores remain more or less unaffected by the choice of a distribution, so consequently the implications of this conclusion are less serious than first expected. Nevertheless, whenever setting a specific distributional assumption in the course of a Stochastic Frontier Analysis, there remains a point open to criticism. Beyond doubt, it would be better to basically abandon unfounded assumptions as far as possible. In this context we showed that the fixed effects approach to inefficiency as proposed by Schmidt and Sickles (1984) overcomes this restriction. Moreover, our adaption of the fixed effects model by Cornwell et al. (1990) to incorporate group-specific trend parameters turned out to be both flexible and robust to estimate.

On page 37 we mentioned another 'red herring' criticism by Førsund et al. (1980) referring to the questionable economic meaning of an average frontier shifted to an extreme. We announced to keep this objection in mind, and, indeed, we found a solution featuring an economic value added: Our proposal to consider Quantile Regression with fixed effects to be a real alternative to the classical SFA turned out to be an innovation in efficiency estimation. Admittedly, we encountered an identification problem between the heterogeneity in the fixed effect and the technical efficiency, but this is by no means the unique feature of Quantile Regression: Greene (2005b) proposed the solution to a 'true' fixed effects model, separating a firm-specific fixed effect from firm-specific inefficiency, thereby coping with exactly the same technical difficulties. Another possible objection regarding our interpretation of the fixed effects to capture the heterogeneity in time refers the fact that this surely affects the resulting time trends of inefficiency. And again, this is not a unique feature of Quantile Regression: On the contrary, it is common practice to incorporate environmental variables into any Stochastic Frontier model (cp. section 5.2). As these control variables contain time trends, too, it is only natural to interpret inefficiency time trends always conditional on environmental time trends that hold for the whole peer group.

References

Afriat, S. N.: 1972, Efficiency estimation and production functions, *International Economic Review* **13**(3), 569–598.

Aigner, D. and Chu, S.: 1968, On the estimation of production frontiers, *American Economic Review* **58**, 826–839.

Aigner, D., Lovell, C. K. and Schmidt, P.: 1977, Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics* **6**, 21–37.

Al-Sharkas, A. A., Hassan, M. K. and Lawrence, S.: 2008, The impact of mergers and acquisitions on the efficiency of the US banking industry: Further evidence, *Journal of Business Finance and Accounting* **36**, 50–70.

Allen, F. and Santomero, A. M.: 1998, The theory of financial intermediation, *Journal of Banking & Finance* **21**, 1461–1485.

Allen, J. and Liu, Y.: 2007, Efficiency and economies of scale of large Canadian banks, *Canadian Journal of Economics* **40**(1), 225–244.

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, B.: 2001, Efficiency in European banking, *European Economic Review* **45**, 1931–1955.

Altunbas, Y., Liu, M.-H., Molyneux, P. and Seth, R.: 2000, Efficiency and risk in Japanese banking, *Journal of Banking & Finance* **24**, 1605–1628.

Andreoni, J.: 1990, Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving, *The Economic Journal* **100**, 464–477.

Bansal, P.: 2005, Evolving sustainably: A longitudinal study of corporate sustainable development, *Strategic Management Journal* **26**(3), 197–218.

Bargain, O. and Kwenda, P.: 2009, The Informal Sector Wage Gap: New Evidence Using Quantile Estimations on Panel Data. Institute for the Study of Labor, Discussion Paper Series, IZA DP No. 4286.

Battese, G. E. and Coelli, T. J.: 1988, Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data, *Journal of Econometrics* **38**, 387–399.

Battese, G. E. and Coelli, T. J.: 1992, Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India, *The Journal of Productivity Analysis* **3**, 153–169.

Battese, G. E., Heshmati, A. and Hjalmarsson, L.: 2000, Efficiency in labour use in Swedish banking industry: a stochastic frontier approach, *Empirical Economics* **25**, 623–640.

Bauer, P. W., Berger, A. N., Ferrier, G. D. and Humphrey, D. B.: 1998, Consistency conditions for regulatory analysis of financial institutions: A comparison of frontier efficiency methods, *Journal of Economics and Business* **50**, 85–114.

Beatti, B. R. and Taylor, C. R.: 1985, *The economics of production*, Wiley, New York.

Behr, A.: 2010, Quantile regression for robust bank efficiency score estimation, *European Journal of Operational Research* **200**(2), 568–581.

Behr, A. and Tente, S.: 2008, Stochastic frontier analysis by means of maximum likelihood and the method of moments. Deutsche Bundesbank Discussion Paper, Series 2, No. 19.

Benston, G. J.: 1965, Branch banking and economies of scale, *The Journal of Finance* **20:2**, 312–331.

Benston, G. J., Hanweck, G. A. and Humphrey, D. B.: 1982, Scale economies in banking: A restructuring and reassessment, *Journal of Money, Credit, and Banking* **14**(4), 435–456.

Berger, A.: 1993, "distribution-free" estimates of efficiency in the u.s. banking industry and test of the standard distributional assumption, *The Journal of Productivity Analysis* **4**, 261–292.

Berger, A. N.: 2007, International comparisons of banking efficiency, *New York University Salomon Center, Financial Markets, Institutions & Instruments* **16**(3), 119–144.

Berger, A. N., Hanweck, G. A. and Humphrey, D. B.: 1987, Competitive viability in banking: Scale, scope, and product mix economies, *Journal of Monetary Economics* **20**, 501–520.

Berger, A. N. and Humphrey, D. B.: 1991, The dominance of inefficiencies over scale and product mix ecnomies in banking, *Journal of Monetary Economics* **28**, 117–148.

Berger, A. N. and Humphrey, D. B.: 1992, Measurement and efficiency issues in commercial banking, *in* Z. Griliches (ed.), *Output Measurement in the Service Sector*, The University of Chicago Press, chapter 7, pp. 245–279.

Berger, A. N. and Humphrey, D. B.: 1997, Efficiency of financial institutions: International survey and directions for future research, *European Journal of Operational Research* **98**, 175–212.

Berger, A. N., Hunter, W. C. and Timme, S. G.: 1993, The efficiency of financial institutions: A review and preview of research past, present, and future, *Journal of Banking and Finance* **17**, 221–249.

Berger, A. N. and Mester, L. J.: 1997, Inside the black box: What explains differences in the effienciencies of financial institutions?, *Journal of Banking & Finance* **21**, 895–947.

Bernini, C., Freo, M. and Gardini, A.: 2004, Quantile estimation of frontier production function, *Empirical Economics* **29**, 373–381.

Berrone, P. and Gomez-Mejia, L. R.: 2009, The pros and cons of rewarding social responsibility at the top, *Human Resource Management* **48**(6), 959–971.

Bos, J. and Koetter, M.: 2007, Handling losses in translog profit models. Munich Personal RePEc Archive No. 1433.

Bos, J. and Kool, C.: 2006, Bank efficiency: The role of bank strategy and local market conditions, *Journal of Banking and Finance* **30**, 1953–1974.

Bos, J. and Schmiedel, H.: 2003, Comparing efficiency in European banking: A meta frontier approach. De Nederlandsche Bank, Research Series Supervision no. 57.

Brachinger, H. W.: 2007, Statistik zwischen Lüge und Wahrheit, *Wirtschafts- und Sozialstatistisches Archiv* **1**, 5–26.

Bundeskriminalamt: 2009, Jahresbericht. Financial Intelligence Unit (FIU - Deutschland).

Burger, A., Frohmüller, K. P. and Moormann, J.: 2008, Produktivität in Banken: warum die CIR (cost income ratio) in die Irre führt, *Bank-Archiv* **56**(12), 860–869.

Canay, I. A.: 2010, A note on Quantile Regression for Panel Data Models. Department of Economics, Northwestern University.

Casu, B. and Girardone, C.: 2005, An analysis of the relevance of off-balance sheet items in explaining productivity change in European banking, *Applied Financial Economics* **15**, 1053–1061.

Casu, B. and Girardone, C.: 2006, Bank competition, concentration and efficiency in the single European market, *The Manchester School* **74**(4), 441–468. Special Issue.

Casu, B. and Girardone, C.: 2009, Testing the relationship between competition and efficiency in banking: A panel data analysis, *Economics Letters* **105**, 134–137.

Casu, B. and Molyneux, P.: 2003, A comparative study of efficiency in European banking, *Applied Economics* **35**, 1865–1876.

Charnes, A., Cooper, W. W. and Rhodes, E.: 1978, Measuring the efficiency of decision making units, *European Journal of Operational Research* **2**, 429–444.

Christensen, L. R., Jorgenson, D. W. and Lau, L. J.: 1973, Transcendental logarithmic production frontiers, *The Review of Economic and Statistics* **55**(1), 28–45.

Clark, J. A. and Speaker, P. J.: 1994, Economies of scale and scope in banking: evidence from a generalized translog cost function, *Quarterly Journal of Business and Economics* **33**, 3–25.

Coelli, T. J.: 1995, Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis, *The Journal of Productivity Analysis* **6**, 247–268.

Coelli, T. J.: 1996, A guide to FRONTIER version 4.1: A computer program for stochastic frontier production and cost function estimation, *Technical Report 7*, Centre for Efficiency and Productivity Analysis (CEPA) Working Papers.

Coelli, T. J., Prasada Rao, D. S., O´Donnell, C. J. and Battese, G. E.: 2005, *An Introduction to Efficiency and Productivity Analysis*, 2 edn, Springer, New York.

Colangelo, A. and Inklaar, R.: 2010, Banking sector output measurement in the EURO area – A modified approach. European Central Bank, Working Paper Series No. 1204.

Colwell, R. J. and Davis, E. P.: 1992, Output and productivity in banking, *Scandinavian Journal of Economics* **94**, 111–129. Supplement.

Cook, R. D. and Weisberg, S.: 1982, *Residuals and Influence in Regression*, Chapman & Hall.

Cornwell, C. and Schmidt, P.: 2008, Stochastic frontier analysis and efficiency estimation, *in* L. Matyas and P. Sevestre (eds), *The Econometrics of Panel Data: Fundamentals and recent developments in theory and practice*, 3 edn, Vol. 46 of *Advanced Studies in Theoretical and Applied Econometrics*, Springer, Berlin, chapter 21, pp. 697–726.

Cornwell, C., Schmidt, P. and Sickles, R. C.: 1990, Production frontiers with cross-sectional and time-series variation in efficiency levels, *Journal of Econometrics* **46**, 185–200.

Cuesta, R. A.: 2000, A production model with firm-specific temporal variation in technical inefficiency: With application to Spanish dairy farms, *Journal of Productivity Analysis* **13**, 139–158.

Demirgüc-Kunt, A. and Detragiache, E.: 2009, Basel Core Principles and Bank Soundness: Does Compliance Matter? Policy Research Working Paper Nr. 5129, The World Bank.

Deutsche Bank AG: 2009, Corporate Social Responsibility Report. **URL:** *www.db.com/csr*

DeYoung, R.: 1997, Measuring Bank Cost Efficiency: Don't count on Accounting Ratios, *Financial Practice and Education* **7**(1), 20–31.

Diamond, D. W.: 1984, Financial intermediation and delegated monitoring, *Review of Economic Studies* **51**(3), 393–414.

Diamond, D. W. and Dybvig, P. H.: 1983, Bank runs, deposit insurance, and liquidity, *The Journal of Political Economy* **91**(3), 401–419.

Dietsch, M. and Lozano-Vivas, A.: 2000, How the environment determines banking efficiency: A comparison between French and Spanish industries, *Journal of Banking & Finance* **24**, 985–1004.

Eichmann, W.: 2005, Finanzserviceleistung, indirekte messung (FISIM). Statistisches Bundesamt, Wirtschaft und Statistik No. 7.

Ennis, H. M.: 2001, On the size distribution of banks, *Federal Reserve Bank of Richmond Economic Quarterly* **87**(4), 1–25.

Farrell, M. J.: 1957, The measurement of productive efficiency, *Journal of the Royal Statistical Society* **120**, 253–281.

Feng, G. and Serletis, A.: 2009, Efficiency and productivity of the US banking industry, 1998-2005: Evidence from the fourier cost function satisfying global regularity conditions, *Journal of Applied Econometrics* **24**, 105–138.

Ferrier, G. D. and Lovell, C. K.: 1990, Measuring cost efficiency in banking, *Journal of Econometrics* **46**, 229–245.

Fitzpatrick, T. and McQuinn, K.: 2005, Cost efficiency in UK and Irish credit institutions, *The Economic and Social Review* **36**(1), 45–66.

Freixas, X. and Rochet, J.-C.: 2008, *Microeconomics of Banking*, second edn, The MIT Press.

Fried, H. O., Lovell, C. A. K. and Schmidt, S. S.: 2008, Efficiency and productivity, *The Measurement of Productive Efficiency and Productivity Growth*, Oxford University Press, New York, pp. 3–91.

Friedman, M.: 1970, The social responsibility of business is to increase its profits, *New York Times Magazin* **13**, 32–33.

Førsund, F. R., Lovell, C. A. K. and Schmidt, P.: 1980, A survey of frontier production functions and of their relationship to efficiency measurement, *Journal of Econometrics* **13**, 5–25.

Girardone, C., Molyneux, P. and Gardener, E. P.: 2004, Analysing the determinants of bank efficiency: the case of Italian banks, *Applied Economics* **36**, 215–227.

Girardone, C., Nankervis, J. C. and Velentza, E.-F.: 2009, Efficiency, ownership and financial structure in European banking: A cross-country comparison, *Managerial Finance* **35**(3), 227–245.

Goddard, J., Molyneux, P. and Wilson, J. O. S.: 2004, The profitability of European banks: A cross-sectional and dynamic panel analysis, *The Manchester School* **72**(3), 363–381.

Greene, W.: 2002, Alternative panel data estimators for stochastic frontier models. Stern School of Business, New York University.

Greene, W.: 2005a, Fixed and random effects in stochastic frontier models, *Journal of Productivity Analysis* **23**, 7–32.

Greene, W.: 2005b, Reconsidering heterogeneity in panel data estimators of the stochastic frontier model, *Journal of Econometrics* **126**, 269–303.

Greene, W. H.: 1980, Maximum likelihood estimation of econometric frontier functions, *Journal of Econometrics* **13**, 27–56.

Greene, W. H.: 1990, A gamma-distributed stochastic frontier model, *Journal of Econometrics* **46**, 141–163.

Greene, W. H.: 1997, Frontier production functions, *in* M. Pesaran and P. Schmidt (eds), *Handbook of applied econometrics*, Blackwell, pp. 81–166.

Greene, W. H.: 2008, The econometric approach to efficiency analysis, *in* H. O. Fried, C. A. K. Lovell and S. S. Schmidt (eds), *The Measurement of Productive Efficiency and Productivity Growth*, Oxford University Press, New York, chapter 2, pp. 92–250.

Grohmann, H.: 1988, Die statistische Adäquation als Postulat einer sachgerechten Abstimmung zwischen Theorie und Empirie, *Theorie und Empirie in der Wirtschaftsforschung*, Freimann, Karsten-Dietmar, pp. 25–42.

Guala, F.: 2008, Evaluating bank efficiency using Cost Frontier and Quantile Regression Analysis: The case of Argentinia in the late 1990s. UIUC Department of Economics Working Paper.

Haldane, A., Brennan, S. and Madouros, V.: 2010, What is the contribution of the financial sector: Miracle or mirage?, *The Future of Finance, The LSE Report*, London School of Economics and Political Science.

Hancock, D.: 1991, *A theory of production for the financial firm*, Kluwer Academic Publishers.

Hao, L. and Naiman, D. Q.: 2007, *Quantile Regression*, number 07-149 in *Quantitative Applications in the Social Sciences*, SAGE Publications.

Hartmann-Wendels, T., Pfingsten, A. and Weber, M.: 2007, *Bankbetriebslehre*, 4 edn, Springer, Berlin.

Hartwig, H.: 1956, Naturwissenschaftliche und sozialwissenschaftliche statistik, *Zeitschrift für die gesamte Staatswissenschaft* **112**, 252–266.

Hjalmarsson, L., Kumbhakar, S. C. and Heshmati, A.: 1996, DEA, DFA and SFA: A comparison, *The Journal of Productivity Analysis* **7**, 303–327.

Hoggarth, G., Reis, R. and Saporta, V.: 2001, Costs of banking system instability: some empirical evidence. Bank of England, Working Papers.

Ioannidis, C., Pasiouras, F. and Zopounidis, C.: 2010, Assessing bank soundness with classification techniques, *Omega* **38**, 345–357.

John, K. and Qian, Y.: 2003, Incentive features in CEO compensation in the banking industry. Federal Reserve Bank of New York, Economic Policy Review.

John, K., Saunders, A. and Senbet, L. W.: 2000, A theory of bank regulation and management compensation, *The Review of Financial Studies* **13**(1), 95–125.

Jondrow, J., Lovell, C., Materov, I. and Schmidt, P.: 1982, On the estimation of technical inefficiency in the stochastic frontier production function model, *Journal of Econometrics* **19**, 233–238.

Kato, K., Galvao, A. F. J. and Montes-Rojas, G. V.: 2010, Asymptotics and bootstrap inference for panel quantile regression models with fixed effects. Working Paper, Graduate School of Science, Department of Mathematics, Hiroshima University.

Keeley, M. C.: 1990, Deposit Insurance, Risk, and Market Power in Banking, *American Economic Review* **80**, 1183–1200.

Kniesner, T. J., Viscusi, W. K. and Ziliak, J. P.: 2009, Policy Relevant Heterogeneity in the Value of Statistical Life: New Evidence from Panel Quantile Regressions. Institute for the Study of Labor, Discussion Paper Series, IZA DP No. 4508.

Koenker, R.: 2004, Quantile regression for longitudinal data, *Journal of Multivariate Analysis* **91**, 74–89.

Koenker, R.: 2005, *Quantile Regression*, Econometric Society Monographs, Cambridge University Press.

Koenker, R. and Bassett, G.: 1978, Regression quantiles, *Econometrica* **46**(1), 33–50.

Koenker, R. and Hallock, K. F.: 2000, Quantile Regression - An Introduction. University of Illinois at Urbana-Champaign.

Koenker, R. and Ng, P.: 2005, A Frisch-Newton Algorithm for Sparse Quantile Regression, *Acta Mathematicae Applicatae Sinica* **21**(2), 225–236.

Koetter, M. and Poghosyan, T.: 2009, The identification of technology regimes in banking: Implications for the market power-fragility nexus, *Journal of Banking and Finance* **33**, 1413–1422.

Kopp, R. J.: 1981, The measurement of productive efficiency: A reconsideration, *The Quarterly Journal of Economics* **96**(3), 477–503.

Kumbhakar, S. C.: 1990, Production frontiers, panel data, and time-varying technical inefficiency, *Journal of Econometrics* **46**, 201–211.

Kumbhakar, S. C. and Lovell, C. K.: 2003, *Stochastic Frontier Analysis*, Cambridge University Press.

Kumbhakar, S. C. and Tsionas, E. G.: 2008, Scale and efficiency measurement using a semiparametric stochastic frontier model: evidence from the U.S. commercial banks, *Empirical Economics* **34**, 585–602.

Lamarche, C.: 2006, Robust penalized Quantile Regression estimation for panel data. Manuscript, Department of Economics, University of Oklahoma.

Lang, G. and Welzel, P.: 1996, Efficiency and technical progress in banking - empirical results for a panel of German cooperative banks, *Journal of Banking & Finance* **20**, 1003–1023.

Leibenstein, H.: 1966, Allocative Efficiency vs. 'X-Efficiency', *American Economic Review* **56**, 392–415.

Litz, H. P.: 1990, Statistische adäquation und idealtypus, anmerkungen zur methodologie der wirtschafts- und sozialstatistik, *Allgmeines Statistisches Archiv* **74**, 429–456.

Liu, C., Laporte, A. and Ferguson, B. S.: 2008, The quantile regression approach to efficiency measurement: Insights from Monte Carlo simulations, *Health economics* **17**(9), 1073–1087.

Lozano-Vivas, A., Pastor, J. T. and Pastor, J. M.: 2002, An efficiency comparison of European banking. Systems operating under different environmental conditions, *Journal of Productivity Analysis* **18**, 59–77.

Macey, J. R. and Miller, G. P.: 1992, Double liability of bank shareholders: History and implications, *Wake Forest Law Review* **27**, 31–62.

Marshall, A.: 1920, *Principles of Economics*, 8 edn, London.

McAllister, P. H. and McManus, D.: 1993, Resolving the scale efficiency puzzle in banking, *Journal of Banking & Finance* **17**, 389–405.

Meeusen, W. and van de Broeck, J.: 1977, Efficiency estimation from Cobb-Douglas production functions with composed errors, *International Economic Review* **18**, 435–444.

Menges, G.: 1982, Die statistische Adäquation, *Jahrbücher für Nationalökonomie und Statistik* **197**(4), 289–307.

Mester, L. J.: 1996, A study of bank efficiency taking into account risk-preferences, *Journal of Banking & Finance* **20**, 1025–1045.

Mlima, A. P. and Hjalmarsson, L.: 2002, Measurement of inputs and outputs in the banking industry, *Tanzanet Journal* **3**, 12–22.

Moormann, J. and Sottocornola, M.: 2009, Die Schattenseiten der Cost-Income-Ratio, *Schweizer Bank* **7**, 24–25.

Oliveira, C. V. and Tabak, B. M.: 2005, An international comparison of banking sectors: A DEA approach, *Global Economic Review* **34**(3), 291–307.

Olson, J. A., Schmidt, P. and Waldman, D. M.: 1980, A Monte Carlo study of estimators of stochastic frontier production functions, *Journal of Econometrics* **13**, 67–82.

Park, S. and Peristiani, S.: 2007, Are bank shareholders enemies of regulators or a potential source of market discipline, *Journal of Banking & Finance* **31**, 2493–2515.

Pasiouras, F.: 2008, International evidence on the impact of regulations and supervision on banks' technical efficiency: an application of two-stage data envelopment analysis, *Review of Quantitative Finance and Accounting* **30**, 187–223.

Pastor, J. M., Pérez, F. and Quesada, J.: 1997, Efficiency analysis in banking firms: An international comparison, *European Journal of Operational Research* **98**, 395–407.

Pastor, J. M. and Serrano, L.: 2006, The effect of specialisation on bank´ efficiency: An international comparison, *International Review of Applied Economics* **20**(1), 125–149.

Perera, S., Skully, M. and Wickramanyake, J.: 2007, Cost efficiency in South Asian banking: The impact of bank size, state ownership and stock exchange listings, *International Review of Finance* **7**(1-2), 35–60.

Pi, L. and Timme, S. G.: 1993, Corporate control and bank efficiency, *Journal of Banking and Finance* **17**, 515–530.

Porter, R. C.: 1961, A model of bank portfolio selection, *Yale Economic Essays* **1**, 323–359.

R Development Core Team: 2010, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing,

Vienna, Austria.
URL: *http://www.r-project.org*

Ritter, C. and Simar, L.: 1997a, Another look at the American electrical utility data. CORE and Institute de Statistique, Université Catholique de Louvain, Working Paper.

Ritter, C. and Simar, L.: 1997b, Pitfalls of normal-gamma stochastic frontier models, *Journal of Productivity Analysis* **8**, 168–182.

Schmidt, P. and Sickles, R. C.: 1984, Production frontiers and panel data, *Journal of Business & Economic Statistics* **2**(4), 367–374.

Sealey, S. and Lindley, J.: 1977, Inputs, outputs and a theory of production and cost at depository financial institutions, *Journal of Finance* **32**, 1251–1266.

Shleifer, A. and Vishny, R. W.: 1997, A survey of Corporate Governance, *The Journal of Finance* **52**(2), 737–783.

Stigler, G. E.: 1976, The Xistence of X-Efficiency, *The American Economic Review* **66**(1), 213–216.

Toby, A. J.: 2006, Methodological approach to the study of X-efficiencies and scale economies in banking: are smaller banks more efficient than larger banks, *Journal of Financial Management and Analysis* **19**(2), 85–96.

Tortosa-Ausina, E.: 2002, Bank cost efficiency and output specification, *Journal of Productivity Analysis* **18**, 199–222.

Weber, M.: 1988, Die 'Objektivität' sozialwissenschaftlicher und sozialpolitischer Erkenntnis, *Gesammelte Aufsätze zur Wissenschaftslehre*, 7th edn, Winckelmann, Johannes.

Weill, L.: 2004, Measuring cost efficiency in European banking: A comparison of frontier techniques, *Journal of Productivity Analysis* **21**, 133–152.

Winsten, C. B.: 1957, Discussion on Mr. Farrell's paper, *Journal of the Royal Statistical Society* **120**, 282–284.

Wooldrigde, J. M.: 2002, *Econometric analysis of cross section and panel data*, The MIT Press.

Lebenslauf