8-2011

Essays on the Performance of US Bank Holding Companies

Liliana Danila
Clemson University, ldanila@clemson.edu

Essays on the Performance of US Bank Holding Companies

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Economics

by
Liliana Danila
August 2011

Accepted by:
Dr. Paul W. Wilson, Committee Chair
Dr. Michael T. Maloney
Dr. Howard Bodenhorn
Dr. Raymond D. Sauer

Abstract

I employ nonparametric estimation techniques to explore different aspects of the performance of US bank holding companies (BHCs).

In my first essay "Market Pricing of U.S. Bank Holding Companies' Technical Efficiency" I use quarterly data from 1986 to 2009 to examine whether changes in technical efficiencies of U.S. BHCs are reflected in their stock returns. The relationship between technical efficiency and stock returns is analyzed both in a financial and accounting framework. Efficiencies are estimated using the unconditional hyperbolic α–quantile estimator developed by Wheelock and Wilson (2008). This estimator is a local estimator and exhibits more desirable statistical properties than traditional estimators. For large bank holding companies, I find evidence of a weak link between technical efficiency and stock returns. I find no other persistent and robust relationship, an indication that the market may not value technical efficiency.

In my second essay, "Restricting the Size of Banks May Have Costs," coauthored with Paul W. Wilson, returns to scale for BHCs are examined. The empirical evidence on scale economies among large BHCs operating in the U.S. is mixed, with some studies finding mild evidence of increasing returns to scale while most studies find no evidence of either increasing or decreasing returns. Most of studies have relied on estimation of fully parametric translog specifications of cost functions. We show that data on BHCs trivially reject the translog specification, and employ fully-

nonparametric methods to estimate and make inference about returns to scale among U.S. BHCs. Our results suggest that both economically and statistically significant increasing returns to scale prevail throughout the range of sizes of BHCs. We use our estimates to provide rough estimates of the cost, in terms of foregone scale economies, of restricting the sizes of BHCs.

My third essay "Evolution of the U.S. Bank Holding Companies' Performance over Time: Evidence from Nonparametric Efficiencies" examines changes in the performance of U.S. BHCs between 1988 and 2010. The Malmquist index measures the total factor productivity change over time and can be decomposed into efficiency change and technology change. I use the nonparametric, unconditional, hyperbolic α–quantile estimator developed by Wheelock and Wilson (2008) to estimate three types of efficiencies: technical, cost, and revenue, that I then use to construct the decompositions of the Malmquist index. Results suggest that over the years, the largest banks experienced the largest gains, on average, in technical, cost and revenue efficiency, with the exception of 2005 -2010 period, and the smallest BHCs seem to have experienced gains in all efficiencies. Estimates of the technology change show a downward shift of theα–quantile (i.e., a decrease in the output produced for some given input used), an upward shift of the cost α–quantile (i.e., an increase in the minimum cost of producing some given output), and downward shift of the revenue α–quantile (i.e., a decrease in the amount of revenue generated from the output produced with some given amount of input), for most periods and class size, except the large BHCs.

Dedication

To my dear family and friends, with love.

Acknowledgments

I wish to thank my committee members, Dr. Paul W. Wilson, Dr. Michael T. Maloney, Dr. Howard Bodenhorn, and Dr. Raymond Sauer, for their advice and continuous support through the process of completing my dissertation work. I would like to especially thank my advisor, Dr. Paul W. Wilson, for his advice, guidance, and patience during my graduate studies. I would not have been able to be where I am today without his unmeasurable support.

I am grateful to Dr. Maloney for his constant support throughout my graduate studies.

I thank Dr. Bodenhorn for his guidance and for the passion for economics and economic history that he instills.

Finally, I am grateful to the faculty members of the John E. Walker Department of Economics. I greatly benefited from their teaching and knowledge. I thank the John E. Walker Department of Economics for the financial support during my graduate school.

Table of Contents

Title Page . . . . **i**

Abstract . . . . **ii**

Dedication . . . . **iv**

Acknowledgments . . . . **v**

List of Tables . . . . **viii**

List of Figures . . . . **ix**

1 Market Pricing of U.S. Bank Holding Companies' Technical Efficiency **1**
1.1 Introduction 1
1.2 Literature Review 7
1.3 Estimation Procedure 10
1.4 Data 18
1.5 Empirical Results 23
1.6 Conclusion 27

2 Restricting the Size of Banks May Have Costs **45**
2.1 Introduction 45
2.2 The Cost Model 50
2.3 Estimation and Inference 55
2.4 Empirical Results 57
2.5 Conclusion 62

3 Evolution of the U.S. Bank Holding Companies' Performance over Time:
Evidence from Nonparametric Efficiencies **75**
3.1 Introduction 75
3.2 Literature Review 78
3.3 Estimation Approach 79

3.4 Data and Results 85
3.5 Conclusion 89

A Appendix: Details of Non–parametric Estimation and Inference . 108
A.1 Dimension Reduction 108
A.2 Estimating Returns to Scale 109

List of Tables

1.1 Inputs and Outputs Definitions 32
1.2 Inputs and Outputs for Different Efficiency Specifications 33
1.3 Pearson Rank Correlation Coefficients of Different Specifications for the Technical Efficiency Scores 34
1.4 Summary Statistics Efficiency Scores 35
1.5 Summary statistics 36
1.6 Pearson Rank Correlation Coefficients between Efficiency and Risk . 38
1.7 Efficiency and Stock Returns: Financial Model 39
1.8 Efficiency and Stock Returns: Accounting Model 40
1.9 Efficiency and Stock Returns: Financial Model, Alternative Efficiency Specifications 41

2.1 Summary statistics 63
2.2 Summary Statistics for Expansion-Path Scale Economy Estimates by Size-Quartile 65
2.3 Expansion–Path Scale Economies, 95–percent Significance 66
2.4 12 Largest Bank Holding Companies as of 30 June 2010 67
2.5 EPSE Estimates for BHCs with Total Assets Exceeding $1 Trillion . . 68

3.1 Number of BHCs over Time 91
3.2 Variables Definitions 92
3.3 Summary Statistics 93
3.4 Number of BHCs in the Sample 95
3.5 Mean Estimates of Efficiency Change by Quartile 96
3.6 Mean Estimates of Technology Change by Quartile 98
3.7 Mean Estimates of Cost Efficiency Change by Quartile 100
3.8 Mean Estimates of Cost Technology Change by Quartile 102
3.9 Revenue Efficiency Change Estimates by Quartile 104
3.10 Revenue Technology Change Estimates by Quartile 106

List of Figures

1.1 Illustration of DEA Estimator 43
1.2 Illustration of α-Quantile Hyperbolic Estimator 44

2.1 Density of (Log) Total Assets for 1986.Q4, 1995.Q4, and 2005.Q4 . . 69
2.2 Ray Scale Economies (D = 0) 70
2.3 Ray Scale Economies (D = 1) 71
2.4 Expansion Path Scale Economies by Size–Quartile, 1986 72
2.5 Expansion Path Scale Economies by Size–Quartile, 1995 73
2.6 Expansion Path Scale Economies by Size–Quartile, 2005 74

Chapter 1

Market Pricing of U.S. Bank Holding Companies' Technical Efficiency

1.1 Introduction

When describing the performance of production units, it is common to characterize them as being (in)efficient, or (non)productive. Productivity refers to the ratio of outputs to inputs, and varies across firms due to differences in technology, production process or production environment. Technical efficiency is one aspect of productivity and it refers to the comparison between observed and potential levels of inputs and outputs, measured by the ratio of observed output from the given level of input to maximum potential output, for the same input level. In other words, technical efficiency refers to the ability to avoid waste by producing the maximum output with the input available (alternatively, it may refer to the ability to avoid waste by using as little input as possible for a given production). Efficiency is technical because it considers solely the production process without taking into account the role of market prices.

The foundations of technical efficiency were set in the 1950's by Koopmans (1951), Debreu (1951), Farrell (1957), and Shephard (1970) who defined different measures of technical efficiency, and by Farrell (1957) and Charnes et al. (1978) who used linear programming techniques for estimation. Since then, the estimation techniques have been refined: from the initial, simple approaches of efficiency point estimates computed from a linear program, today statistical methods are available for statistical inference of technical efficiency.[1] The body of literature in this field has grown considerably over the past thirty years, with applications to a large number of industries particularly banking, mutual funds, aviation, nursing, hospitals, agriculture, transportation, electricity, and country-level studies such as the U.S., Australia, Canada, Belgium, Greece, Italy, China, Norway, Israel. Gattoufi et al. (2004) provide a comprehensive survey of the literature that examines technical and allocative efficiency. They cite approximately 2,000 studies published in no fewer than 490 distinct refereed journals published worldwide between 1951 and 2001. More than thousand studies have been published since.

Given academics' wide interest in technical efficiency as a performance measure, it is interesting to investigate whether investors value it and to what extent (i.e. does a relationship between technical efficiency and stock returns exist?). This paper examines whether changes in the quarterly technical efficiency of U.S. bank holding companies (BHCs) are reflected in their quarterly stock returns. BHCs are analyzed rather than subsidiary banks, since managers presumably make decisions with the goals of the entire institution in mind.

[1] See Cooper et al. (2004) and Simar and Wilson (2000) for a review of past and current inference methods.

The question of a causal relationship between technical efficiency and stock returns is an empirical one and no theoretical prediction of the relationship between the two variables exists. I expect, however, that a positive and significant association exists based on the fact that future profits and operating efficiency are potentially interconnected. Financial theory holds that a company chooses positive net present value projects to maximize shareholders' wealth. Technical efficiency is an indicator of current operational performance. Higher efficiency translates either into higher output produced with the same inputs, or the same output level with less inputs. If relative prices of both inputs and outputs are unchanged, then producing more output with the same inputs leads to an increase in profitability. Also, even if technical efficiency does not imply profit efficiency, technically inefficient firms are necessarily not profit-efficient. Assuming constant prices of inputs and outputs, if two similar firms use the same amount of inputs, and one produces less output, it has to be the case that it is not maximizing its profits, relative to the other one. Thus, an improvement in current performance should result, ceteris paribus, in higher future profits. If the market values future profitability, a change in technical efficiency should be incorporated in stock returns.

Conversely, Modigliani and Miller (1958) recognize that when a production process is characterized by uncertainty, the firm's objective is value maximization, rather than profit maximization, because the latter does not account for production risk or the appropriate discount rate applied to the profit stream.[2] Given the profit volatility in the highly leveraged banking industry, in addition to profitability, investors examine a multitude of sources of information in search of indicators of BHCs' performance.

[2] Under uncertainty, there corresponds to each decision of the firm not a unique profit outcome, but a plurality of mutually exclusive outcomes which can at best be described by a subjective probability distribution.

This paper offers two innovations to the existing literature on banking efficiency. First, it estimates technical efficiency employing the new non-parametric unconditional hyperbolic α-quantile estimator developed by Wheelock and Wilson (2008). The α-quantile estimator exhibits superior statistical properties in comparison to the traditional nonparametric efficiency estimators, such as data envelopment analysis or free disposal hull. In contrast to the traditional estimators, this is a local estimator: the benchmark comparison is given only by firms similar to the one analyzed, and not the by the entire sample. This is particularly significant when analyzing the banking industry. The U.S. banking industry consists of different size BHCs where small BHCs with only few million dollars in assets operate side by side with extremely large ones that have trillions of dollars in assets. The local α-quantile estimator ensures a relevant relative comparison base as it enforces comparison of firms of the same size.

The second innovation of this research comes from the richness of the data set used. In contrast to the majority of studies in the field that use cross-sectional data, my data consist of a sample of bank holding companies covering a period of more than twenty years (1986 – 2009). There are over 1,000 BHCs in almost all of the 91 quarters analyzed, with a maximum of 1,830 observations in one quarter. This number of observations is considerably higher than the sample size used in the majority of the existing studies that analyze, on average, a cross section of few hundred observations.

The nature of the banking data in relation to technical efficiency has to be discussed before proceeding with the analysis. The only BHC data source available, to my knowledge, is provided by the quarterly reports that BHCs have to submit, by law, to the Fed. While this is a rich source of information, the empirical banking field recognizes the quarterly reports exhibit some drawbacks when it comes to estimating efficiency in a frontier estimation framework. One of them relates to the information available on the total loan deposit amount only, without the number of accounts. As other authors recognize too, it is not possible to distinguish output quality and higher output implies higher efficiency. Loans made represent a bank's output, but writing more loans does not necessarily imply a higher efficiency level. Likewise, the book value of loans is assumed to be the same as their market value, leading to incorrect conclusions since excessive loan growth will be reflected as high efficiency.[3] Foos et al. (2009) examine the intertemporal relationship between abnormal loan growth and the riskiness of banks on a sample of 16,000 European banks between 1997-2007. They find that loan growth leads to an increase in loan loss provisions during the next three years, a decrease in relative interest income, and to lower capital ratios. These findings suggest that loan growth drives the riskiness of banks.

Other feature of the banking data is that market prices are not accounted for. Thus, estimated efficiencies are technical: they account for the technical aspect of the production of using too many inputs or producing too few outputs. The technical efficiency does not reflect the allocative aspect of efficiency of (mis)responding to relative prices in choosing inputs or outputs, or the management aspect of engaging in high–risk activities. Since the banking industry is a highly leveraged, accounting for risk is extremely important for a true description of the banking firm. But a BHC that is very efficient, by technical efficiency measures, may also be very risky, since it operates at "too high" leverage levels. In an attempt to mitigate the effect of lack of market prices, I follow Hughes et al. (2001) and include the BHC's capital as an input in the production function. Hughes et al. (2001) document that BHCs' scale economies are uncovered when equity capital, in addition to debt, is included in the production model.

[3] In many banks, higher loan volume is the result of misaligned incentives: employees are rewarded for the number of accounts they open, without regard for the quality of the loans given out.

Finally, some may be concerned with the fact that accounting data are used to estimate efficiencies, as accounting data are past–looking, while stock prices are forward-looking. However, I do not consider the accounting data an issue. From a theroretical perspective, the price of a firms share in the market equals the expected discounted payoff that does not depend on past performance. Financial theory holds that markets are efficient and securities are fairly priced, with three levels of market efficiency defined, distinguished by the amount of information incorporated in security prices.[4] Under the semi-strong version of the efficient market hypothesis, accounting information is incorporated in stock prices. Since news about a change in efficiency of a firm is public information, it is captured in its return, and thus there should be a relation between the two.

Empirically, a consistent robust finding in the research examining the link between accounting information and stock performance is that investors do react to earnings announcements regardless of the industry or the time frame analyzed. Patell and Wolfson (1984) found that the market adjusts very quickly when a firm publishes its latest earnings or announces a dividend change. A major adjustment in price occurs within five to ten minutes of the announcement. On the other hand, Bernard and Thomas (1989) documented that investors underreact to the earnings announcement and become aware of its full significance only as further information arrives: it takes about 60 days until the market fully incorporates the new information, regardless of the news being good or bad.

[4]The three efficiency levels are: (i) *the weak form* of efficiency: prices reflect the information contained in the record of past prices. If markets are efficient in the weak sense, then it is impossible to make consistently superior profits by studying past returns; (ii) *the semistrong form* of efficiency: prices reflect not just past prices, but all other published information, such as the information in the financial press or official governmental statistics. If markets are efficient in this sense, then prices will adjust immediately to public information such as the announcement of the last quarter's earnings; and, (iii) *the strong form* of efficiency: prices reflect all the information that can be acquired by painstaking analysis of the company and the economy.

Practically, players in the financial markets are expected to evaluate a firm's performance based on a multitude of sources and indicators. Given the U.S. free market system and free information flow technical efficiency may be one of the sources of information for investors. Bognini et al. (2002) document the performance of three sets of indicators: accounting data, stock prices, and credit ratings in forecasting financial distress for banks in East Asian countries (1996–1998). They find that even if the stock market has responded quicker, it did not significantly outpace accounting information contained in the balance sheet or assessments by the credit agencies.[5] If the market has the ability to process effectively the information available and if efficiency scores capture some information relevant to the market, then there should be an association between technical efficiency and stock returns.

The remainder of this paper is structured as follows: Section 1.2 presents the literature review, Section 1.3 describes the estimation procedure, Section 1.4 describes the data, Section 1.5 presents the results, and Section 1.6 concludes the paper.

1.2 Literature Review

Berger and Humphrey (1997) surveyed 130 studies of efficiency analysis of financial institutions in 21 countries. They find, unsurprisingly, that various efficiency models and estimators yield different results depending on the frontier used and on how output is measured (flow or stock variable), with slightly higher estimates from nonparametric studies. Once a model and an output specification are adopted, the

[5] On the other hand, the current financial crisis showed that there are times when the market does not uncover a bubble while it's developing. The official report on Lehman Brothers (2010) showed that the company had used Repo 105 (purchase and resale) transactions to remove approximately $50 billion of liquid assets from the balance sheet in 2008 in order to decrease its net leverage. This information was not seized by the market. Secretary Geithner stated: "If this had been a bank we were supervising, that [i.e., Lehmans Repo 105 program] would have been a huge issue for the New York Fed." Thus, revealed accounting statements, may reveal information not available otherwise.

estimates are fairly stable.

While research on efficiency in the financial/ banking sector is very rich, there are only a handful of papers that analyze the association between stock markets and technical efficiency. Eisenbeis et al. (1999) analyze "the informativeness of stochastic frontier and programming frontier" by examining a sample of 254 BHCs over the time period 1986-1991. For their sample, they find that while both the stochastic and the programming frontier measures result in a similar ranking of firms' efficiencies, estimates derived from the stochastic frontier are associated with risk-taking behavior, managerial performance, and stock returns. Efficiency estimated nonparametrically does not exhibit such a pattern (it shows little consistent association with any of the above measures). The authors conclude that the programming efficiencies are not economically meaningful, thus they are not informative, as apposed to the stochastic ones.

Beccalli and Casu (2006) analyze a sample of European publicly traded banks in the year 2000. They regress the annual stock returns on annual change in efficiency, estimated by both stochastic frontier analysis (SFA) and data envelopment analysis (DEA), proxies for size, risk and profit, and country dummies. Their results suggest that 1 percent change in the annual efficiency score derived from DEA leads to an approximately 0.4 percent change in annual bank returns (or a 0.2 percent change, if performance proxies and country dummies are included in the model). The SFA estimates show a similar pattern, but less strong. The inclusion of accounting data to proxy for riskiness or profitability does not seem to increase the explanatory power of the model.

Results obtained by Kirkwood and Nahm (2006) similarly suggest that changes in firms' efficiencies are reflected in stock returns. They examine a sample of 67 Australian banks over the 1995-2002 time period using the three-factor capital asset pricing model. The change in profit efficiency is positive and statistically significant for the regional banks, but insignificant for the major banks. However, the authors interpret the results as an indication of positive association between changes in efficiency and returns. The beta value for the major banks was estimated to be 1.43, and beta for the regional banks is 0.58, suggesting, against the general belief, that major banks returns are more sensitive to overall market movements than the regional banks returns.

Ioannidis et al. (2008) estimate the stochastic cost and profit efficiency of a sample of Asian and Latin American publicly traded banks over the 2000–2006 period. They regress the annual bank returns on annual efficiencies changes, and find a positive, robust relationship between the profit efficiency and the stock returns, but no evidence of such a relationship for cost efficiency. The authors' explanation for this finding is that shareholders are more interested in their wealth, thus are more interested in dividend payments and capital gains. Since dividends will be paid on the basis of profits, stock returns are more sensitive to profit efficiency rather than cost efficiency.

Another way to think about whether the market should value technical efficiency is to examine whether this measure can predict future outcomes. Wheelock and Wilson (1995) offer evidence in this sense. They analyze the failures of commercial banks in Kansas in early twentieth century. Among the factors examined to explain bank failure, they use a measure of technical efficiency estimated from a stochastic frontier. Modelling time-to-failure in a proportional hazards framework, and find that, among other determinants, technically inefficient banks were more likely to fail than technically efficient banks.

Apart from the banking industry, Semenick Alam and Sickles (1998) examined the connection between stock prices and technical efficiency using U.S. airline data for a sample of eleven companies for a twenty-year period. A quarterly efficiency score calculated for each firm employing the data envelopment analysis (DEA), and the free disposal hull (FDH) is compared with the return. They employ a simple correlation test; to check for robustness, they construct portfolios of the top and bottom firms and verify that the average return for the top portfolio is consistently larger. Their results show that there is a statistically and economically significant relationship between technical efficiency changes and market returns. The correlation is observed in the two months after the financial data is disclosed.

1.3 Estimation Procedure

1.3.1 Nonparametric Technical Efficiency Estimation

The nonparametric approaches of analyzing efficiency, generally known as data envelopment techniques, are "data oriented" techniques for estimating the performance of an economic agent relative to its peer agents by analyzing the production process of transforming the inputs available into outputs. Formally, data envelopment techniques are "methodologies directed to frontiers rather than central tendencies" (Cooper et al., 2004). As opposed to trying to fit a regression plane through the center of the data, in DEA set-up, one "floats" a piecewise linear surface to rest on top of the observations. Both stochastic and nonparametric techniques exist for estimation of technical efficiency.

DEA methods have the main advantage of not having to specify a functional relationship between variables a priori. This offers a great level of flexibility: few assumptions are necessary and, in particular, no assumptions about the frontier shape or the distribution of inputs and outputs on the production set are necessary. The most popular nonparametric estimators are the free disposal hull (FDH) and data envelopment analysis, and are based on the idea of estimating the attainable set by enveloping the data. Charnes et al. (1978) operationalized the concept of efficiency in terms of linear programming, such that efficiency in the presence of multiple input and output factors is defined as the ratio between the weighted sum of outputs and the weighted sum of inputs, and each firm selects input and output weights to maximize its efficiency score. A "deterministic frontier" based on the observed data is estimated, and using linear programs, the efficiency scores are estimated. These efficiency measures are relative to the other data observed. Thus, full efficiency is attained by any firm if and only if none of its inputs or outputs can be improved without worsening some of its other inputs or outputs; given the deterministic frontier, there will be observations that lie on the frontier (i.e., they are the most efficient in either input or output direction).

Efficiencies estimated by DEA or FDH use the full frontier as a benchmark, and indicate the minimum achievable level of input (or the maximum achievable level of output) over all production plans that are technically feasible. The frontier is sometimes referred to in the literature as "the observed best practice frontier," since observations that lie on it are the economic agents the most efficient (they use the least amount of inputs and obtain the highest amount of output). While DEA or FDH estimators exhibit several advantages over the parametric estimators, a trade-off exist: no functional relationship has to be specified for the nonparametric estimators, thus no "risk" of misspecification is encountered, as opposed to the parametric estimators. But the parametric estimators have a high convergence rate, namely $n^{(-1/2)}$. Kneip et al. (1998) show that the convergence rate for the DEA estimator is $n^{-1/(p+q)}$ and $n^{-2/(p+q)}$ for the FDH (where p refers to the number of inputs used, and q to the number of outputs used). As the number of inputs and outputs used increases, the size of the sample has to increase almost exponentially for the DEA or FDH estimators to convergence, hence the "curse of dimensionality" for the nonparametric estimators. Another shortcoming of the traditional estimators is that they are consistent only asymptotically. Though bootstrapping offers an alternative to estimation, the curse of dimensionality is still present: the sample size has to be quite large for the estimators to be consistent. Finally, the DEA and FDH estimators are extremely sensitive to outliers.

1.3.2 Hyperbolic α-quantile Estimation

Wheelock and Wilson (2008) extended results obtained by Daouia and Simar (2007) and developed the unconditional hyperbolic α-quantile estimator in order to deal with the drawbacks of traditional nonparametric estimators. The α-quantile estimator is robust to outliers in the data and has a high convergence rate (comparable to the parametric estimators' convergence rate). Besides the better statistical properties, there are other features that make this estimator a more desirable one in practice. First, the α-quantile estimator is a "partial" frontier estimator: the benchmark comparison is not the entire population of firms (as it is the case for traditional estimators), but only the ones similar to the firm analyzed. The BHCs analyzed in this paper differ greatly in size: although all firms operate in the same industry, their operations are not comparable. The α-quantile estimator is a local estimator, in the sense that it allows comparison of firms in the neighborhood of the analyzed firm. Even more, the researcher can choose the desired size of the neighborhood (i.e. one can choose to use 1 percent of the sample observations as benchmark or 10 percent).

The second desirable feature of the α-quantile is that it allows estimation of the efficiency along a hyperbolic path, such that inputs and outputs are adjusted simultaneously, rather than just in the input or output direction, as it happens with traditional estimators. In reality, when intending to change the scale of operations, it is only rarely that one can act only in the input or in the output direction. The traditional estimators estimate the distance from a fixed data point to the full frontier in a direction orthogonal to the output axis (in the input orientation case), or to the input axis (in the output orientation case). If the sector analyzed operates under variable returns to scale, then the choice of input or output orientation has a big impact on the measured efficiency: a large firm could lie close to the frontier in the output direction, but far from it in the input direction, while a small firm could be close to the frontier in the input direction, but far from in the output direction. By allowing estimation of the efficiency along a hyperbolic path where inputs and outputs are adjusted simultaneously, the α-quantile estimator overcomes this issue.

Figures 1.1 and 1.2 illustrate these concepts, where it is assumed that the production process is characterized by a single- input and single- output (assumption made only for simplification, so that the process can be illustrated in a two-dimensional graph). Figure 1.1 shows the DEA approach for measuring efficiency. The deterministic, piecewise-linear DEA frontier is shown, as determined by the data. There are few observations: A, B, and C, representing individual firms, that lie inside the production set, and not on the production frontier. A comparison between agents A and B based on the input distance function reveals that A is significantly more efficient than B: an efficiency estimate in the input direction would be found by estimating the distance from the individual firms to the frontier in a direction orthogonal to y. We can see that A lies much closer to the frontier in the input direction, than B does. However, if efficiency in the output direction is estimated, then the opposite is found: B is more efficient than A, since B lies closer to the frontier in the output direction.[6] Analyzing the agents A and B, the natural implication is that they operate at different production scales, and thus a comparison between the two of them would not be extremely revealing. The α-quantile estimation method, illustrated in Figure 1.2, deals with this issue by estimating the distance to the frontier in the hyperbolic direction, instead of the orthogonal one. If agent A is the observation of interest, then its efficiency is given by the estimate distance to the frontier estimate in the hyperbolic path. This is more meaningful in an economic sense because it is more reasonable to assume that alteration of the inputs mix would result in an change on the outputs side too, and vice versa. Also, this method allows the researcher to set the size of the sample against which firms should be benchmarked. In the exampled shown in figure 2, if A is to be compared to 30 percent of the sample in the neighborhood of A, then the efficiency estimate is the estimation of the distance from A to the frontier along the hyperbolic path. If A is to be compared to only 15 percent, then the efficiency estimate is the estimation of the distance from the projection of A, the point represented in the figure by A', to the frontier along the hyperbolic path.

Formally, we can describe the model in the following manner. This presentation is based on Wheelock and Wilson (2008). Consider the following production possibilities set:

$$P \equiv \{(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x} \text{ can produce } \boldsymbol{y}\} \subset \mathbb{R}_{+}^{p+q}, \tag{1.1}$$

where $\boldsymbol{x} \in \mathbb{R}_{+}^{p}$ and $\boldsymbol{y} \in \mathbb{R}_{+}^{q}$ denote vectors of inputs and outputs, respectively, and P^{δ} denotes the upper boundary of the production set P, referred to as the technology frontier.

The goal is to estimate the distance from an observation, a point $(\boldsymbol{x}, \boldsymbol{y})$ in the space P, to the frontier P^{δ}.

[6] There are agents, like the one at C, for which measuring efficiency in the input or output direction does not seem to make a difference.

The hyperbolic-graph distance function measures the distance from a fixed point $(\boldsymbol{x}, \boldsymbol{y})$ to P^δ along the hyperbolic path: $(\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y})$, where $\gamma \subset \mathbb{R}^1_{++}$. The hyperbolic distance function is given by

$$\gamma(\boldsymbol{x}, \boldsymbol{y} \mid P) \equiv \sup\left\{\gamma > 0 \mid (\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y}) \in P\right\}. \tag{1.2}$$

In order to measure the efficiency along this hyperbolic path, we have to first estimate the frontier. Note that P represents the true production set, out of which we observe only a sample. The data sample observed comprises realizations of iid random variables with probability density function $f(\boldsymbol{x}, \boldsymbol{y})$ with support over P. Let $(\boldsymbol{x}_0^\delta, \boldsymbol{y}_0^\delta) \in P^\delta$ denote a point of the frontier P^δ. The density $f(\boldsymbol{x}, \boldsymbol{y})$ implies a probability function that gives the likelihood of drawing an observation from the $f(\boldsymbol{x}, \boldsymbol{y})$ that weakly dominates the agent operating at $(\boldsymbol{x}_0^\delta, \boldsymbol{y}_0^\delta) \in P^\delta$. An observation $f(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$ weakly dominates f($\boldsymbol{x}$,$\boldsymbol{y}$) if $\widetilde{\boldsymbol{x}} \leq \boldsymbol{x}$ and $\widetilde{\boldsymbol{y}} \geq \boldsymbol{y}$. Formally, we express this probability function in the following way:

$$H(\boldsymbol{x}_0, \boldsymbol{y}_0) = Pr(\boldsymbol{x} \leq \boldsymbol{x}_0, \boldsymbol{y} \geq \boldsymbol{y}_0). \tag{1.3}$$

Using the probability function defined above, we can express the hyperbolic distance function as

$$\gamma(\boldsymbol{x}, \boldsymbol{y} \mid P) = \sup\left\{\gamma > 0 \mid H(\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y}) > 0\right\} \tag{1.4}$$

and the hyperbolic α-quantile distance function as

$$\gamma_\alpha(\boldsymbol{x}, \boldsymbol{y}) = \sup\left\{\gamma > 0 \mid H(\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y}) > (1-\alpha)\right\}. \tag{1.5}$$

For $0 < \alpha < 1$ and a fixed point $(\boldsymbol{x}, \boldsymbol{y}) \in \Re_{+}^{p+q}$, $\gamma_\alpha(\boldsymbol{x}, \boldsymbol{y}) > 1$ gives the proportionate, simultaneous decrease in inputs and increase in outputs required to move from $(\boldsymbol{x}, \boldsymbol{y})$ along a path $(\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y})$, $\gamma > 0$, to a point with $(1-\alpha)$ probability of being weakly dominated. The hyperbolic α-quantile frontier is defined by:

$$P_\alpha^\delta = \left\{ (\gamma_\alpha(\boldsymbol{x}, \boldsymbol{y})^{-1}\boldsymbol{x}, \gamma_\alpha(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{y}) \mid (\boldsymbol{x}, \boldsymbol{y}) \in P \right\}. \tag{1.6}$$

To estimate the hyperbolic α-quantile distance function and frontier, the probability function defined in (3.3) is replaced by its empirical analog. Using the bisection method, the estimator of the distance function, $\widehat{\gamma}_\alpha(\boldsymbol{x}, \boldsymbol{y})$, can be computed. Wheelock and Wilson (2008) expose at length the estimation procedure.

1.3.3 Bank Efficiency and Stock Performance

To examine the causal relationship between the two measures of firm's performance, the technical efficiency and the stock returns, I estimate two empirical models. One model makes use of the financial framework. I estimate the below model that is routinely used in the literature to examine the influence of a "nontraditional" factor on stock returns, where changes in the S&P 500 return and changes in the technical efficiency estimates are related to changes in the BHCs' returns. Considering a firm's valuation in the market depends on the overall macroeconomic conditions (changes in interest rates, economic growth, industry performance) and on firm's specifics (firm's ability to generate future cash flows, its capital structure, its ability to "innovate" by undertaking projects that give a higher rate of return to the cost of capital), the S&P 500 return is included to reflect the overall changes in the market conditions, and the change in technical efficiency to reflect the overall changes at the firm level. The model to be estimated is presented below, and time fixed effects, or/and quartile dummies, or/and interaction terms between efficiency scores and quartiles.[7]

$$R_{it} = \beta_0 + \beta_1 SP500_t + \beta_2 \Delta\widehat{\gamma}_\alpha(\boldsymbol{x}, \boldsymbol{y})_{it} + \epsilon_{it}, \tag{1.7}$$

where R_{it} designates the quarterly return of BHC i in quarter t, $SP500_t$ the S&P 500 return in quarter t, $\Delta\widehat{\gamma}_\alpha(\boldsymbol{x}, \boldsymbol{y})_{it}$ is the estimated change in the α-quantile hyperbolic estimate of technical efficiency for the BHC i in quarter t, and ϵ_{it} is a normally distributed error term. Since the banking data is reported quarterly, in order to have full correspondence between the two measures, quarterly returns were calculated. The change in estimated efficiency was calculated as a percentage change between the efficiency estimates of two consecutive quarters: $\Delta\widehat{\gamma}_\alpha(\boldsymbol{x}, \boldsymbol{y})_{it} = [(\widehat{\gamma}_\alpha(\boldsymbol{x}, \boldsymbol{y})_{it} - \widehat{\gamma}_\alpha(\boldsymbol{x}, \boldsymbol{y})_{it-1})/\widehat{\gamma}_\alpha(\boldsymbol{x}, \boldsymbol{y})_{it-1}]$.

The second model uses the accounting set-up. To date, extensive research has failed to find strong links between accounting data and stock returns. Zhang (2000) and Chen and Zhang (2006) provide a theoretical model and an empirical test to explain cross-sectional variation in the stock returns by accounting data. Their valuation model is based on the the fact that firm value consists of the value of owned assets plus the value of the growth opportunities (thus maintaining that stock returns are forward looking and reflect future profitability). Valuation amounts to forecasting the scale and profitability of future operations. Returns are estimated as a function of earnings yield, equity capital investment, changes in profitability, in growth opportunities and in the discount rate. Their model explains about 20 percent of returns variation in their sample that comprises firm-level data covering the period 1983-2001.

I build on the model developed by Zhang (2000) and Chen and Zhang (2006),

[7] I divided each quarterly data into four quartiles based on their assets size. The fourth quartile contains the largest BHCs.

with the following differences. First, I use the change in estimated efficiency score as a proxy for change in profitability. Second, the authors employ the consensus analyst forecasts of long-term earning growth from the I/B/E/S database as a proxy for the growth opportunities. I employ the ratio of loan provisions to total loans to proxy for the growth opportunities. This ratio indicates the reserves banks set aside to cover potential future losses in the loan portfolio. The higher this ratio is, the more loans are forecasted to be charged-off in the future. Hence, the higher this ratio is, the lower is the future growth potential of the bank. I estimate the following model, also employing time fixed effects, or/and quartile dummies, or/and interaction terms between efficiency scores and quartiles:

$$R_{it} = \alpha_0 + \alpha_1 \Delta\widehat{\gamma}_\alpha(\boldsymbol{x}, \boldsymbol{y})_{it} + \alpha_2 Earn_{it} + \alpha_3 Eq_{it} + \alpha_4 Growth_{it} + \alpha_5 r_t + e_{it}, \quad (1.8)$$

where R_{it} and $\Delta\widehat{\gamma}_\alpha(\boldsymbol{x}, \boldsymbol{y})_{it}$ are defined in the same way as in the financial model. $Earn_{it}$ shows the quarterly change in earnings of BHC i in quarter t, Eq_{it} is the quarterly change in the book value of equity of BHC i in quarter t, $Growth_{it}$ is loan provision to total loans ratio of BHC i in quarter t, and r_t is the change in the discount rate in quarter t relative to quarter $t - 1$. As with the change in efficiency, the changes in different variables were calculated as percentage changes between the values of two consecutive quarters.

1.4 Data

The data on BHCs are collected from the Federal Reserve Bank's (FBR) FR-Y 9C forms (FRB of Chicago website), which contain information on BHCs required to report their accounting statements to the Fed. The sample comprises data on all

BHCs that reported to the Fed, between 1986:Q4 - 2009:Q2. Starting March 2006, the Fed increased the asset-size threshold for filing the FR Y-9C form from $150 million to $500 million. The stock return data were collected from the Center in Research in Security Prices (CRSP). Also, the CRSP-FRB link data set, made available by the New York Fed, was used to merge the two data sets.[8] The GDP deflator data was collected from the St. Louis Fed website.

1.4.1 Input /Output Specification

The specification of inputs and outputs of bank production is part of an ongoing debate. I use the intermediation approach, according to which banking institutions buy and sell funds, thus they act as intermediaries between borrowers and lenders. In this framework, deposits are considered inputs and loans and investments outputs. I employ the intermediation approach motivated not solely by data availability, but also following Sealey and Lindley (1977). Sealey and Lindley (1977) develop a model of behavior of financial firms consistent with the neoclassical theory of the firm. They attempt to classify inputs and outputs of the financial firm by considering the technical aspect of production for the financial firm. Their production model is defined, following Frisch (1965), as a"transformation" process, where "certain goods and/or services (inputs) enter into a process in which they loose their identity, i.e. cease to exist in the original form, while other goods and services (outputs) are generated.[9]

Table 1.1 shows the definitions of inputs and outputs chosen for the technical

[8]The CRSP-FRB data are avalaible at:
http://www.newyorkfed.org/research/banking_research/datasets.html.

[9]Three approaches to modeling production functions exist in the literature: (i) the intermediation approach, (ii) the intermediation approach, according to which funds are collected from depositors and banks compete to attract depositors (deposits are considered outputs in this set-up), and (iii) the user cost approach, that defines inputs and outputs based on their contribution to revenue.

score specification that I use as the main specification in this paper. As inputs, I use labor, physical capital, core deposits, purchased funds, equity, charged-off loans (used as the ratio of 1 to charged-off loans, to reflect the "non desirable" nature of the charged-off loans) and recovered charged-off loans. I include equity as an input since equity may be used as an alternative to deposits in funding loans. This may account, to a certain extend, for the management of risk. Furthermore, Mester (1996) argues that the inclusion of equity in the analysis may account for differences in bank managers risk attitude, since higher levels of equity reduce the risk of default. Berger and Mester (1997) argue that equity should be included, given the greater dependence of large banks on debt financing than smaller banks.

I include the charged-off and recovered loans in an attempt to account for unmeasured differences in output quality.[10] Berger and Mester (1997) try to mitigate the same problem by including the ratio of charged-off loans to total loans in the state. Concerned that nonperforming loans may be endogenous to the production function, they chose to use the ratio of nonperforming loans rather than just the level of non-performing loans. In theory, nonperforming loans are exogenous if they are caused by negative economic shocks. Berger and Mester (1997) argue, however, that if we consider the production model, all output variables are ultimately endogenous since they are chosen by the bank. Considering this argument, and the fact that banks do not choose per se the level of nonperforming loans, I consider that the nonperforming loans are exogenous to the production function and, thus, it is appropriate to include this variable as an input in the production specification.

As outputs, I use securities, real estate loans, consumer loans, business loans,

[10]The reason why I include two variables, rather than just one: charged-off loans net of recoveries, is that there are BHCs that do not have any charged-off loans in some quarters, but they do have recovered loans. Thus, if I included only one variable, the resulting input would have some negative values.

other loans (calculated as the difference between total loans and the summation of the loan categories) and off-balance items. The reason for defining different loan categories as different outputs, rather than just one output, total loans, is to allow for some heterogeneity among firms. Off-balance activities are included since, recently, banks have increasingly engaged in more non-traditional banking activities.

Table 1.2 shows other specifications used to estimate the efficiency scores, as robustness checks: the "basic" specification does not consider the charged-off loans and recovered loans as inputs, or the off-balance activities as output. It only includes inputs and outputs describing the traditional banking activities, thus "basic" specification. The other specification, referred to as the "inclusive" specification, differs from the "basic" one by considering the off-balance activities as an output.

I estimated efficiency scores using the three different specifications, and different sample size for the comparison groups: 1 percent of the sample (corresponding to a value of α of 0.99), 5 percent (α=0.95), and 10 percent (α=0.90). Table 1.3 presents the Pearson correlation coefficients between these different efficiency measures for the entire sample and selected quarters. The correlation coefficient between the "preferred" specification and the "inclusive" one is around 0.80 for similar benchmark groups, for the entire sample of just quarterly data; similarly, the correlation between the "preferred" specification and the "basic" one is approximately 0.75, and between the "inclusive" and the "basic" one of approximately 0.90. Thus, we can conclude they are relatively similar (with the "inclusive" and the "basic" specification being very similar). The reason that the favorite specification was the "preferred" is based on the attempt to allow for firm heterogeneity and account for risk to a certain extent, and thus it is desirable to account for nonperforming loans.

Table 1.4 presents the summary statistics for the efficiency scores. The FEAR package by Wilson (2008) was used to obtain the efficiency estimates. Given the difference in the firms, detailed statistics are presented for each bank quartile: banks were divided into four quartiles based on asset size in each quarter. Comparison of the efficiency score across years is not informative since it is a relative measure: it relates the new performance of the BHC to performance of BHCs in the benchmark, that may have experienced changes themselves. Within the same quarter, however, firms in the second quartile are the most inefficient (the high score translates into low efficiency), in almost each of the quarters analyzed. It does not appear that there is a consistent significant difference between the average efficiency of a small BHC versus a large one. On the other hand, given small BHCs operate at different scale than the large BHCs, the comparison may not have economic significance. In the first quarter of 2007, the average estimated efficiency for large banks was 0.4638. That is, on average they were using about 46.38 percent of the input amount and produced more than double (1/0.4638) the output of a BHC located on the frontier of the comparison group. Also, it seems that in each quartile, smaller banks are more efficient than their larger counterparts.

Table 1.5 presents the summary statistics for input and output variables for the overall sample and for selective quarters (all values in constant 2005 \$). The striking feature of the data is the significant difference in BHC size: in the first quarter of 1997, while the the average value for total assets was $\$4,432,190,000$, the average value for total assets for the smallest quartile (by asset size) was only $\$229,804,000$. There is a similar difference between the BHCs in other respects too: total loans, liabilities, deposits,... It is important to emphasize that a partial frontier efficiency score is even more desirable in this context: given the wide distribution of BHCs, it is obvious that full frontier estimates would not be meaningful. The average size of a bank increased over time, and also the difference between large and small banks increased too. The fact that the number of BHCs increased overtime (by almost 20 percent between 1987 and 1997) is not surprising. The BHC legal status offers some advantages over the "bank" only status, i.e. it is easier to borrow from the Fed. Also, in 1994 the Riegle-Neal Interstate Banking and Branching Efficiency Act that allowed banks to operate in different states was passed. In 1999, the Gramm-Leach-Bliley Act was passed, allowing BHCs to engage in non-traditional banking activities (i.e.: financial activities, insurance underwriting). This act was an incentive for more banks to adopt the BHC status. However, the lower number of banks in 2007 reflects the change in the reporting rules imposed by the Fed in 2006.

1.5 Empirical Results

A priori, if the financial markets are efficient and if the nonparametric efficiency measure captures inefficiencies relevant to the market, we would expect a relationship between the α-quantile estimator and the stock returns. Aware of the fact that, by its nature, the banking business implies a high level of risk[11], and that the efficiency measure may capture some risk by including equity as an input, Table 1.6 presents the Pearson correlation coefficients between the α-quantile estimator and the accounting measure of risk: the ratio of equity to total assets. This measure reflects the capital adequacy or financial leverage of the company. According to Eisenbeis et al. (1999), this ratio also captures the degree to which shareholders have their own capital at risk in the institution, and thus may reflect their incentives to monitor management and assure that the institution operates efficiently. The degree of correlation between this ratio and the efficiency is quite low. Examining the entire sample, the correlation is positive for the first three quarters, implying that a more risky bank is more inefficient too (by construction, a *lower value of the α-quantile estimator* is indicative of *higher*

[11]Agents in the stock market attempt not only to account for risk, but they analyze the different types of banking risks: solvency, liquidity, credit, interest rate, and operating risk, separately.

efficiency). For the largest banks, the coefficient shows an opposite correlation: the largest banks are more risky and more efficient. However, the correlation relationship does not appear to be stable: each quartile exhibits negative or positive correlation in different quarters. Although correlation does not imply causation, and the equity to assets ratio is only a broad measure of risk, we can conclude that the efficiency measure does not appear to be reflective of risk.

Table 1.7 shows the results from the financial model shown in (3.7), where the quarterly stock returns are regressed against the quarterly change in efficiency, the S&P 500 return, with or without quartile dummies, time fixed effects, or interaction terms. The efficiency scores were estimated using all the available data (110,642 observations), but the regression analysis includes only BHCs that are publicly traded, hence the sample reduces to 27,328 observations. A negative coefficient for the efficiency change variable would be indicative of a negative association between *inefficiency* and returns. Column (1) of Table 1.7 is an attempt to show whether a "raw" relationship between the variables of interest exists: the change in efficiency coefficient, though it has the expected sign, has very low magnitude and is not significant. Specifications shown in columns (2) – (4), where the S&P500 return, year fixed effects, and quartile dummies where included, indicate that the change in efficiency does not have a causal implication for the stock return.

If the size of the BHC has an impact on the operations of the company (given the non-normality of the inputs/output data), then interaction terms between efficiency and quartile should be included. The results from these specifications are shown in column (5) and (6), the latter including time effects too. Except for the large banks, results do not show an association between the α-quantile estimators and the stock returns: as in the previous regressions, the coefficient of interest has a very low value, is insignificant, and does not exhibit the same sign in different regressions.

The returns of large banks seem to reflect the change in efficiency: if a BHC in the fourth quartile improves its efficiency by 1 percent, then its quarter stock return will increase by 0.013 percent, or by 0.008 percent if year effects are added to the regression. The significance level and the sign of the coefficients could be interpreted in this case as weak evidence of a causal relationship: maybe the large banks are riskier and their operations have a higher degree of complexity, such that the market agents analyze them more carefully. If we take into account, however, the magnitude of the coefficient, and if we corroborate the findings that the market does reflect changes in efficiency for the banks of other size, it is difficult to claim the weak evidence wholeheartedly. Another surprising finding is that the size of the bank does not seem to have an impact on the holding period return. As expected, consistent with the theory and prior research in the financial field, the coefficient on the S&P500 return is positive and strongly significant in all regressions.

The results in Table 1.7 employ the "preferred" specification for the efficiency score, estimated against a benchmark of 1 percent of the sample. As a robustness check, the same model was analyzed with different input/output specifications for the efficiency score or different benchmark sample sizes. Results are shown in Table 1.9. The first panel employs an efficiency estimator without the charged-off loans and recovered charged-off loans as inputs, and the benchmark sample size of 1 percent. In the second panel, charged-off loans and recovered charged-off loans were excluded from the inputs category, and the off-balance items from the outputs for the efficiency score calculation, with a base comparison group of 1 percent of the sample. The results differ in some respects from the main specification. It appears that there is a negative association between improvement in efficiency and stock returns for the smallest BHCs (belonging to the first quartile). The weak link between efficiency change and market returns for the large BHCs is not confirmed: significance level differs, and the estimated effect varies by estimation (it has both positive and negative values).

In the third and fourth panel of Table 1.7, the "preferred" specification is employed, but the comparison group is increased to 5 percent and 10 percent, respectively, of the sample. The causality between change in technical efficiency and stock returns for large BHCs is confirmed by these estimations. Thus, if the "preferred" specification is used, regardless of the sample size employed as a benchmark, then it seems there is an association between the change in efficiency and stock returns. This suggests that accounting for charged-off loans is important.

The results from the accounting model from (1.7) are presented in Table 1.8. Similar with the financial model outcome, the efficiency coefficients are significant for the large banks, and also for the BHCs in the second quartile, but have a low magnitude: 1 percent increase in efficiency causes a 0.015 percent change in the quarter returns of the large banks, or 0.029 percent for the second quartile banks. The ratio of loan provisions to total loans exhibit a strong and robust causal relationship with the stock return. The market does not appear to react to a change in the ten-year discount rate and the change in the quarterly earnings.

Thus, it could be concluded that the above analysis shows weak evidence of a positive causal relationship between stock returns and change efficiency estimated by the nonparametric α-quantile estimator. The results are in line with those obtained by Eisenbeis et al. (1999), even if they used a much smaller sample and a different method for estimating the technical efficiency.

1.6 Conclusion

During recent decades, the research on firms technical efficiency has grown at a significant pace. In this paper, I combine the capital market research and the efficiency literature, by examining the relationship between this measure of performance and the stock returns, the market performance measure, for a sample of BHCs that ranges between 1986 and 2009. The unconditional hyperbolic α-quantile estimator was used to estimate the efficiency score. This estimator, developed by Wheelock and Wilson (2008), has superior statistical properties relative to other existing estimators: it is robust to outliers in the data, has a high convergence rate, and allows for simultaneous inputs and outputs adjustment.

The data provide weak evidence of a causal relationship between technical efficiency and stock returns for the large banks. But no other consistent, robust relationship is found, regardless of the econometric specification used, in the financial or the accounting set-up, or of the efficiency score specification. I interpret this as indication that investors do not value technical efficiency. Possible explanations for this finding may be that the efficiency scores do not incorporate risk (a defining feature of both stock returns and banking activity). Also, they do not account for output quality, such that more loans (outputs) are associated, ceteris paribus, with higher efficiency. Although I included equity, charged-off loans and recovered-charged-off loans in the efficiency score specification in an attempt to control for these factors, the outcomes were unchanged.

Bibliography

Amess, K. & S. Girma (2009), 'Do stock markets value efficiency', *Scottish Journal of Political Economy* **56**(3), 321–331.

Beccalli, E., B. Casu & C. Girardone (2006), 'Efficiency and stock performance in European banking', *Journal of Banking, Finance and Accounting* **33**(1, 2), 245–262.

Berger, A. N. & L. J. Mester (1997), 'Inside the black box: What explain differences in the efficiencies of financial institutions', *Journal of Banking and Finance* **21**(7), 895–947.

Bernard, V. L. & J. K. Thomas (1989), 'Post-earnings-announcement drift: Delayed price response or risk premium?', *Journal of Accounting Research* **27**, 1–36.

Bongini, P., L. Laeven & G. Majnoni (2002), 'How good is the market at assessing bank fragility? A horce race between different indicators', *Journal of Banking and Finance* **26**, 1011–1028.

Charnes, A., W. W. Cooper & E. Rhodes (1978), 'Measuring the efficiency of decision making units', *European Journal of Operational Research* **2**, 429–444.

Chen, P. & G. Zhang (2007), 'How do accounting variables explain stock price movements? Theory and evidence', *Journal of Accounting and Economics* **43**, 219–244.

Cooper, W. W., L. M. Seiford & J. Zhu (2004*a*), Data envelopment analysis: History, models, and interpretations, *in* W. W.Cooper, L. M.Seiford & J.Zhu, eds, 'Handbook on Data Envelopment Analysis', Kluwer Academic Publishers, pp. 1–40.

Cooper, W. W., L. M. Seiford & J. Zhu (2004*b*), *Handbook on Data Envelopment Analysis*, Kluwer Academic Publishers.

Daouia, A. & L. Simar (2007), 'Nonparametric efficiency analysis: A multivariate conditional quantile approach', *Journal of Econometrics* **140**, 375–400.

Debreu, G. (1951), 'The coefficient of resource utilization', *Econometrica* **19**, 273–292.

Eisenberg, R. A., G. D. Ferrier & S. H. Kwan (1999), The informativeness of stochastic frontier and programming frontier efficiency scores: Cost efficiency and other measures of bank holding company performance. Working Paper #99-23, Federal Reserve Bank of Atlanta.

Fama, E. F. & K. R. French (2004), 'The capital asset pricing model: Theory and evidence', *The Journal of Economic Perspectives* **18**(3), 25–46.

Farrell, M. J. (1957), 'The measurement of productive efficiency', *Journal of Royal Statistical Society* **120**, 253–281.

Foos, D., L. Norden & M. Weber (2009), Loan growth and riskiness of banks. SSRN Working Paper #1045001.

Gattoufi, S., M. Oralb & A. Reisman (2004), 'Data envelopment analysis literature: a bibliography update (19512001)', *Socio-Economic Planning Science* **38**, 159–229.

Haddad, M. D., M. J. B. Hall, K. Kenjegalieva, W. Santoso, R. Satria & R. Simper (2008), Banking efficiency and stock market performance: An analysis of listed Indonesian banks. Discussion Paper Series, Loughborough University, Department of Economics.

Hughes, J. P. (1999), 'Incorporating risk into the analysis of production', *Atlantic Economic Journal* **27**(1), 1–23.

Hughes, J. P., L. J. Mester & C.-G. Moon (2001), 'Are scale economies in banking elusive or illusive? Evidence obtained by incorporating capital structure and risk-taking into models of bank production', *Journal of Banking and Finance* **25**(12), 2169–2208.

Ioannidis, C., P. Molyneux & F. Pasiouras (2008), The relationship between bank efficiency and stock returns: Evidence from Asia and Latin America. Working Paper Series, University of Bath, School of Management.

Kirkwood, J. & D. Nahm (2006), 'Australian banking efficiency and its relation to stock returns', *The Economic Record* **82**(258), 253–267.

Kneip, A., B. U. Park & L. Simar (1998), 'A note on the convergence of nonparametric DEA estimators for production efficiency scores', *Econometric Theory* **14**, 783–793.

Koopmans, T. J. (1951), Analysis of production as an efficient combination of activities, *in* T. C.Koopmans, ed., 'Activity Analysis of Production and Allocation', John Wiley & Sons, New York, pp. 33–97.

Laeven, L. (1999), Risk and efficiency in East Asian banks. World Bank Policy Research Working Paper #2255.

Mester, L. J. (1997), 'Measuring efficiency at U.S. banks: Accounting for heterogeneity is important', *European Journal of Operational Research* **98**(2), 230–242.

Modigliani, F. & M. H. Miller (1958), 'The cost of capital, corporation finance and the theory of investment', *American Economic Review* **48**(3), 261–297.

Patell, J. & M. Wolfson (1984), 'The intraday speed of adjustment of stock prices to earnings and dividend announcements', *Journal of Financial Economics* **13**, 223–252.

Shephard, R. W. (1970), *Theory of Cost and Production Functions*, Princeton University Press, Princeton, NJ.

Simar, L. & P. W. Wilson (2000), 'Statistical inference in nonparametric frontier models: The state of the art', *Journal of Productivity Analysis* **13**, 49–78.

Wheelock, D.C. & P. W. Wilson (1995), 'Explaining bank failures: Deposit insurance, regulation, and efficiency', *The Review of Economics and Statistics* **77**(4), 689–700.

Wheelock, D.C. & P. W. Wilson (2008), 'Non-parametric, unconditional quantile estimation for efficiency analysis with an application to Federal Reserve check processing operations', *Journal of Econometrics* (145), 209–225.

Wilson, P. W. (2008), 'Fear 1.0: A software package for frontier efficiency analysis with R', *Socio-Economic Planning Sciences* (42), 247–254.

Zhang, G. (2000), 'Accounting information, capital investment decisions, and equity valuation: Theory and empirical implications', *Journal of Accounting Research* **38**(2), 271–295.

Table 1.1: Inputs and Outputs Definitions

	Inputs
Labor	number of full-time equivalent employees
Physical capital	premises and fixed assets (including capitalized leases)
Equity	total bank holding company equity capital plus noncontrolling (minority) interests in consolidated subsidiaries
Core deposits	domestic transactions accounts, time deposits under \$100,000, and savings deposits
Purchased funds	time deposits over \$100,000, Federal funds purchased in domestic offices, securities sold under agreements to repurchase, trading liabilities, other borrowed money, subordinated notes and debentures, subordinated notes payable to unconsolidated trusts issuing trust preferred securities, and trust preferred securities issued by consolidated special purpose entities
Charged-off loans	
Recovered charged-off loans	
	Outputs
Securities	held-to-maturity securities, available-for-sale securities, federal funds sold in domestic offices, securities purchased under agreements to resell
Real estate loans	loans secured by real estate
Consumer loans	loans to individuals for household, family, and other personal expenditures
Business loans	commercial and industrial loans
Other loans	
Off-balance items	commercial and similar letters of credit, securities lent, financial standby letters of credit conveyed to others, performance standby letters of credit and foreign office guarantees, unused commitments (revolving, open-end loans secured by 1−4 family residential properties, credit card lines, commitments to fund commercial real estate, construction, and land development loans secured by real estate), all other off-balance sheet items

Table 1.2: Inputs and Outputs for Different Efficiency Specifications

	Preferred specification	**Inclusive specification**	**Basic specification**
Inputs	Labor	Labor	Labor
	Physical capital	Physical capital	Physical capital
	Equity	Equity	Equity
	Core deposits	Core deposits	Core deposits
	Purchased funds	Purchased funds	Purchased funds
	Charged-off loans		
	Recovered charged-off loans		
Outputs	Securities	Securities	Securities
	Real estate loans	Real estate loans	Real estate loans
	Consumer loans	Consumer loans	Consumer loans
	Business loans	Business loans	Business loans
	Other loans	Other loans	Other loans
	Off-balance items	Off-balance items	

Table 1.3: Pearson Rank Correlation Coefficients of Different Specifications for the Technical Efficiency Scores

	pref. α**=.99**	**pref.** α**=.95**	**pref.** α**=.90**	**incl.** α**=.99**	**incl.** α**=.95**	**incl.** α**=.90**	**basic** α**=.99**	**basic** α**=.95**
				Entire Sample (n=110,663)				
pref., α**=.99**	1							
pref., α**=.95**	0.9353	1						
pref., α**=.90**	0.8739	0.982	1					
incl., α**=.99**	0.8355	0.7952	0.7456	1				
incl., α**=.95**	0.7847	0.8527	0.8458	0.9292	1			
incl., α**=.90**	0.7296	0.8381	0.8601	0.8645	0.9814	1		
basic, α**=.99**	0.7918	0.7694	0.7265	0.9405	0.8962	0.8408	1	
basic, α**=.95**	0.7444	0.8211	0.8186	0.8799	0.9615	0.9498	0.9327	1
basic, α**=.90**	0.6953	0.8086	0.8322	0.8229	0.9475	0.9681	0.8742	0.9832
				1987 Q1 (n=980)				
pref., α**=.99**	1							
pref., α**=.95**	0.9152	1						
pref., α**=.90**	0.7955	0.9606	1					
incl., α**=.99**	0.8278	0.7694	0.6811	1				
incl., α**=.95**	0.7509	0.8364	0.8324	0.9101	1			
incl., α**=.90**	0.6359	0.7995	0.8616	0.7892	0.9625	1		
basic, α**=.99**	0.7357	0.7013	0.6344	0.8865	0.8278	0.7316	1	
basic, α**=.95**	0.6696	0.7734	0.7877	0.8117	0.9212	0.9057	0.903	1
basic, α**=.90**	0.5748	0.7486	0.8192	0.7093	0.895	0.9428	0.7931	0.967
				1997 Q1 (n=1,170)				
pref., α**=.99**	1							
pref., α**=.95**	0.9241	1						
pref., α**=.90**	0.8552	0.9814	1					
incl., α**=.99**	0.8374	0.8088	0.7615	1				
incl., α**=.95**	0.7734	0.8667	0.8661	0.927	1			
incl., α**=.90**	0.7144	0.8525	0.8800	0.8629	0.9829	1		
basic, α**=.99**	0.7860	0.7770	0.7391	0.9353	0.8900	0.8374	1	
basic, α**=.95**	0.7277	0.8310	0.8376	0.8754	0.9614	0.9537	0.9296	1
basic, α**=.90**	0.6776	0.8213	0.8520	0.8200	0.9496	0.9706	0.8714	0.9850
				2007 Q1 (n=819)				
pref., α**=.99**	1							
pref., α**=.95**	0.9455	1						
pref., α**=.90**	0.9073	0.9882	1					
incl., α**=.99**	0.8154	0.7854	0.7484	1				
incl., α**=.95**	0.7834	0.8349	0.8224	0.9387	1			
incl., α**=.90**	0.7506	0.8247	0.8270	0.8963	0.9873	1		
basic, α**=.99**	0.7912	0.775	0.7418	0.9615	0.9252	0.8883	1	
basic, α**=.95**	0.7621	0.8180	0.8085	0.9080	0.9766	0.9686	0.9447	1
basic, α**=.90**	0.7319	0.8080	0.8121	0.8695	0.9660	0.9791	0.9063	0.9887

Table 1.4: Summary Statistics Efficiency Scores

	Quartile1	**Quartile 2**	**Quartile 3**	**Quartile 4**
Entire Sample (n=110,663)				
Mean	0.4763	0.4983	0.4653	0.3863
St. Dev.	0.1160	0.1060	0.1068	0.1244
Q1	0.4014	0.4335	0.4765	0.3132
Median	0.4871	0.5068	0.5414	0.4021
Q3	0.5596	0.5742	0.5935	0.4765
1987 Q1 (n=980)				
Mean	0.4074	0.4869	0.4537	0.4296
St. Dev.	0.1073	0.0888	0.1126	0.1146
Q1	0.3333	0.4293	0.3879	0.3654
Median	0.4029	0.4887	0.4735	0.4471
Q3	0.4781	0.5500	0.5331	0.5113
1997 Q1 (n=1,170)				
Mean	0.5084	0.5076	0.4780	0.3963
St. Dev.	0.1196	0.0893	0.1038	0.1126
Q1	0.4286	0.4519	0.4248	0.3333
Median	0.5245	0.5095	0.4888	0.4122
Q3	0.5906	0.5731	0.5488	0.4818
2007 Q1 (n=819)				
Mean	0.4847	0.5015	0.4638	0.4638
St. Dev.	0.1245	0.1157	0.1110	0.1110
Q1	0.4137	0.4416	0.3955	0.2963
Median	0.4933	0.5195	0.4749	0.3865
Q3	0.5792	0.5822	0.5411	0.4554

Table 1.5: Summary statistics

	Mean	**Std. Dev.**	**Q1**	**Median**	**Q3**
Entire Sample (n=110,663)					
Total assets	6,100,083	51,900,000	242,078	433,550	1,129,811
Total costs	104,833	799,108	4,116	7,297	18,624
Profits	23,084	269,337	765	1,512	3,904
Total loans	3,608,477	25,500,000	157,398	288,911	767,426
Total liabilities	5,827,420	47,900,000	236,272	418,234	1,070,686
Noninterest income	37,487	363,674	493	1,045	3,186
Inputs					
Labor	1,881	11,645	116	196	452
Physical capital	76,390	431,716	4,613	8,822	21,668
Equity	596,200	4,517,289	27,316	46,845	118,658
Core deposits	2,744,793	16,500,000	190,410	325,417	785,939
Purchased funds	2,078,081	22,200,000	36,233	77,435	248,070
Charged-off loans	10,550	114,828	66	217	905
Recovered charged-off loans	1,839	17,168	17	52	195
Outputs					
Securities	1,392,112	12,300,000	66,807	120,677	299,324
Real estate loans	1,709,302	12,000,000	92,728	183,955	495,855
Consumer loans	644,172	5,842,661	12,466	26,667	76,493
Business loans	881,427	6,101,187	22,321	47,868	130,993
Other loans	383,835	3,010,831	3,232	11,129	40,809
Off-balance items	1,103,365	15,400,000	2,582	10,069	38,014
1987 Q1 (n=980)					
Total assets	3,795,049	15,800,000	221,296	410,894	1,403,011
Total costs	78,466	379,143	4,472	8,176	28,001
Profits	9,030	37,401	369	1,091	3,683
Total loans	2,393,417	10,300,00	112,926	230,664	848,387
Total liabilities	3,568,604	15,000,000	206,275	376,947	1,285,257
Noninterest income	14,912	83,195	382	917	3,629
Inputs					
Labor	2,212	8,445	143	282	980
Physical capital	61,312	242,712	3,658	8,535	26,448
Equity	356,229	1,286,615	20,727	49,227	165,226
Core deposits	1,875,426	5,556,574	163,229	317,685	1,043,141
Purchased funds	926,761	4,192,237	23,648	59,891	243,918
Charged-off loans	5,805	36,411	85	305	1,160
Recovered charged-off loans	1,208	8,215	27	85	330
Outputs					
Securities	664,681	2,051,295	58,785	123,775	365,965
Real estate loans	712,615	3,102,376	43,770	99,657	315,671
Consumer loans	444,544	2,128,172	18,807	44,900	181,751
Business loans	845,569	3,785,074	26,340	63,796	233,453
Other loans	416,791	2,145,176	6,245	18,777	79,332
Off-balance items	189,484	2,344,651	161	753	4,256

Table 1.5 – continued

	Mean	Std. Dev.	Q1	Median	Q3
	1997 Q1 (n=1,170)				
Total assets	4,432,190	21,600,000	229,804	376,840	1,073,498
Total costs	90,266	449,070	4,460	7,389	20,270
Profits	26,408	126,920	1,140	1,976	5,604
Total loans	3,183,299	14,900,000	165,806	274,968	782,309
Total liabilities	4,858,634	23,900,000	247,783	407,373	1,168,022
Noninterest income	28,551	159,345	502	1,029	3,229
Inputs					
Labor	1,823	7,947	129	214	529
Physical capital	80,447	380,173	4,985	8,969	23,360
Equity	547,289	2,545,608	29,810	49,761	152,147
Core deposits	2,695,989	10,600,000	203,383	331,915	916,502
Purchased funds	1,461,441	9,489,714	34,511	64,516	228,719
Charged-off loans	6,442	37,156	57	168	734
Recovered charged-off loans	1,731	10,294	21	58	252
Outputs					
Securities	1,125,085	5,477,960	82,310	139,664	358,544
Real estate loans	1,366,521	5,730,347	100,762	172,701	478,783
Consumer loans	624,501	3,294,704	17,407	37,415	113,188
Business loans	860,930	4,662,541	23,543	47,747	139,589
Other loans	337,574	1,917,014	3,157	10,508	40,626
Off-balance items	810,107	5,419,348	4,341	12,275	38,249
	2007 Q1 (n=819)				
Total assets	13,700,000	109,000,000	612,550	938,739	2,123,228
Total costs	203,459	1,634,139	9,006	13,957	29,803
Profits	55,278	463,326	1,683	3,011	7,506
Total loans	6,765,415	45,300,000	410,898	659,278	1,428,176
Total liabilities	12,000,000	96,300,000	532,036	803,658	1,839,683
Noninterest income	80,540	753,449	1,181	2,191	5,548
Inputs					
Labor	2,308	16,912	173	258	534
Physical capital	109,704	595,826	10,360	18,760	35,225
Equity	1,019,335	7,446,113	46,812	72,895	163,980
Core deposits	4,578,369	27,900,000	378,471	563,919	1,175,876
Purchased funds	5,682,955	52,200,000	141,655	240,250	563,203
Charged-off loans	13,294	136,245	63	203	721
Recovered charged-off loans	2,808	26,111	21	72	238
Outputs					
Securities	2,994,788	26,700,000	98,690	181,032	405,115
Real estate loans	3,873,845	23,700,000	305,075	499,047	1,067,459
Consumer loans	1,047,298	9,727,503	10,173	24,996	66,496
Business loans	1,303,287	8,798,100	49,073	93,405	215,434
Other loans	546,462	4,642,534	4,403	15,994	61,426
Off-balance items	2,955,181	32,300,000	22,866	46,898	125,203

Table 1.6: Pearson Rank Correlation Coefficients between Efficiency and Risk

	Entire Sample (n=110,642)	**1987:Q1** (n=979)	**1997:Q1** (n=1,170)	**2007:Q1** (n=819)
Quartile 1	0.0902	0.1628	0.0822	-0.0668
Quartile 2	0.0295	0.0820	-0.0007	0.1092
Quartile 3	0.0208	0.0145	0.0199	-0.1096
Quartile 4	-0.0871	0.2453	0.0665	0.1878

Table 1.7: Efficiency and Stock Returns: Financial Model

	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x}, \boldsymbol{y})$	-0.0004	-0.0008	-0.0001	-0.0008	-0.008	-0.012
	(0.001)	(0.001)	(0.001)	(0.001)	(0.020)	(0.019)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x}, \boldsymbol{y})$*2nd quartile					-0.009	-0.010
					(0.012)	(0.011)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x}, \boldsymbol{y})$*3rd quartile					0.0002	0.001
					(0.001)	(0.001)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x}, \boldsymbol{y})$*4th quartile					-0.013**	-0.008*
					(0.005)	(0.005)
S&P return		0.554***	0.551***	0.554***	0.554***	0.551***
		(0.012)	(0.014)	(0.012)	(0.012)	(0.014)
2nd quartile				0.003	0.003	-0.001
				(0.006)	(0.006)	(0.006)
3rd quartile				0.009*	0.009	-0.001
				(0.005)	(0.006)	(0.005)
4th quartile				0.009	0.009	0.001
				(0.006)	(0.006)	(0.005)
constant	0.035***	0.024***	-0.071***	0.058***	0.016***	-0.072***
	(0.011)	(0.001)	(0.012)	(0.005)	(0.005)	(0.013)
n=27,328						

NOTES:
1. Specification (3) includes time fixed effects, Specification (4) includes quartile dummies, Specification (5) includes quartile dummies and interaction terms, and Specification (6) includes quartile dummies, interaction terms, and time fixed effects.
2. Results of panel ordinary least squares regression of the effects of a change in the contemporaneous change in efficiency on stock returns.
3. Standard errors in parentheses; *** indicates significance at 1% level, ** at 5% level, and * at 10% level, respectively.
4. Coefficients for the interaction terms show the point estimate of the effect of a change in efficiency of a BHC belonging to a specific quartile on stock return.
5. $\Delta\widehat{\gamma}_\alpha(\boldsymbol{x}, \boldsymbol{y})$ is the quarterly change in technical efficiency.
6. The quartile variables are dummy variables that control for size: in each quarter, BHCs were divided in four size quartiles, where the first quartile contains the smallest banks, and the fourth one the largest BHCs. The benchmark is the first quartile. Time fixed are year fixed effects.

Table 1.8: Efficiency and Stock Returns: Accounting Model

	(1)	(2)	(3)	(4)	(5)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$	-0.001	-0.0006	-0.001	-0.027	-0.023
	(0.001)	(0.001)	(0.001)	(0.021)	(0.020)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$*2nd quartile				-0.029**	-0.028**
				(0.012)	(0.011)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$*3rd quartile				0.001	0.001
				(0.001)	(0.001)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$*4th quartile				-0.022***	-0.015***
				(0.005)	(0.004)
change in earnings	0.0001	0.0001	0.0001	0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
change in equity	0.011**	0.007	0.011**	0.011**	0.007
	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)
loan prov./total loans	-9.192***	-7.190***	-9.202***	-9.318***	-7.280***
	(0.365)	(0.348)	(0.365)	(0.366)	(0.348)
change in discount rate	0.0001	0.0001	0.0001	0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
2nd quartile			0.003	0.003	-0.003
			(0.007)	(0.007)	(0.006)
3rd quartile			0.013*	0.012*	-0.003
			(0.007)	(0.007)	(0.006)
4th quartile			0.010	0.010	0.002
			(0.007)	(0.007)	(0.006)
constant	0.048***	-0.001	0.039***	0.040***	-0.001
	(0.002)	(0.006)	(0.007)	(0.007)	(0.008)
n=27,328					

NOTES:
1. Specification (2) includes time fixed effects, Specification (3) includes quartile dummies, Specification (4) includes quartile dummies and interaction terms, and Specification (5) includes quartile dummies, interaction terms, and time fixed effects.
2. Standard errors in parentheses; *** indicates significance at 1% level, ** at 5% level, and * at 10% level, respectively. Coefficients for the interaction terms show the point estimate of the effect of a change in efficiency of a BHC belonging to a specific quartile on stock return. The benchmark is the first quartile (the smallest BHCs).
3. $\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$ is the quarterly change in technical efficiency.
4. Time fixed are year fixed effects.

Table 1.9: Efficiency and Stock Returns: Financial Model, Alternative Efficiency Specifications

	(1)	(2)	(3)	(4)	(5)	(6)
Inclusive efficiency specification, α=.99						
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$	0.0004	-0.0001	0.0005	-0.0001	0.116**	0.104**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.051)	(0.048)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$*2nd quartile					0.012	0.007
					(0.030)	(0.029)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$*3rd quartile					0.0002	0.001
					(0.001)	(0.001)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$*4th quartile					-0.019**	-0.008
					(0.009)	(0.008)
S&P return		0.554***	0.551***	0.554***	0.554***	0.551***
		(0.012)	(0.014)	(0.012)	(0.012)	(0.014)
2nd quartile				0.003	0.004	-0.001
				(0.006)	(0.006)	(0.005)
3rd quartile				0.009*	0.010*	-0.0002
				(0.005)	(0.006)	(0.005)
4th quartile				0.009	0.009*	0.002
				(0.006)	(0.006)	(0.005)
constant	0.035***	0.024***	-0.071***	0.058***	0.016***	-0.072***
	(0.001)	(0.001)	(0.012)	(0.005)	(0.005)	(0.013)
n=27,328						
Basic efficiency specification, α=.99						
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$	0.0008	0.0002	0.0006	0.0002	0.131**	0.116**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.055)	(0.052)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$*2nd quartile					0.015	0.007
					(0.036)	(0.034)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$*3rd quartile					0.0001	0.0005
					(0.001)	(0.0001)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$*4th quartile					0.005	0.026*
					(0.015)	(0.014)
S&P return		0.554***	0.511***	0.554***	0.554***	0.551***
		(0.012)	(0.014)	(0.012)	(0.012)	(0.014)
2nd quartile				0.003	0.004	-0.001
				(0.006)	(0.006)	(0.006)
3rd quartile				0.009*	0.010	-0.0004
				(0.005)	(0.006)	(0.005)
4th quartile				0.009	0.009*	0.002
				(0.006)	(0.005)	(0.005)
constant	0.035***	0.024***	-0.071***	0.058***	0.016***	-0.072***
	(0.014)	(0.001)	(0.012)	(0.005)	(0.005)	(0.013)
n=27,328						

Table 1.9 – continued

	(1)	(2)	(3)	(4)	(5)	(6)
Preferred efficiency specification, α=.95						
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$	-0.001	-0.001	-0.001	-0.001	-0.012	-0.013
	(0.001)	(0.001)	(0.001)	(0.001)	(0.014)	(0.013)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$*2nd quartile					-0.006	-0.005
					(0.008)	(0.008)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$*3rd quartile					0.0001	0.0005
					(0.001)	(0.0009)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$*4th quartile					-0.011***	-0.006*
					(0.003)	(0.003)
S&P return		0.554***	0.5518***	0.554***	0.553***	0.550***
		(0.012)	(0.014)	(0.012)	(0.012)	(0.014)
2nd quartile				0.003	-0.114***	-0.083**
				(0.006)	(0.030)	(0.030)
3rd quartile				0.009*	0.009	-0.001
				(0.006)	(0.006)	(0.005)
4th quartile				0.009	0.009	0.0001
				(0.006)	(0.006)	(0.005)
constant	0.035***	0.024***	-0.071***	0.058***	0.016***	-0.707***
	(0.001)	(0.001)	(0.012)	(0.005)	(0.005)	(0.013)
n=27,386						
Preferred efficiency specification, α=.90						
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$	-0.001	-0.001	-0.0003	-0.001	-0.011	-0.012
	(0.0008)	(0.0008)	(0.0007)	(0.0008)	(0.011)	(0.010)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$*2nd quartile					-0.005	-0.003
					(0.007)	(0.006)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$*3rd quartile					0.0001	0.0004
					(0.0009)	(0.0008)
$\Delta\widehat{\gamma}_\alpha(\boldsymbol{x},\boldsymbol{y})$*4th quartile					-0.009***	-0.004**
					(0.002)	(0.002)
S&P return		0.554***	0.511***	0.554***	0.554***	0.550***
		(0.012)	(0.014)	(0.012)	(0.012)	(0.014)
2nd quartile				0.003	0.003	-0.002
				(0.006)	(0.006)	(0.006)
3rd quartile				0.009*	0.009	-0.002
				(0.006)	(0.006)	(0.005)
4th quartile				0.009	0.008	0.0004
				(0.006)	(0.006)	(0.005)
constant	0.035***	0.024***	-0.071***	0.058***	0.016***	-0.071***
	(0.014)	(0.001)	(0.012)	(0.005)	(0.005)	(0.013)
n=27,328						

NOTE: Specification (3) includes time fixed effects, Specification (4) includes quartile dummies, Specification (5) includes quartile dummies and interaction terms, and Specification (6) includes quartile dummies, interaction terms, and time fixed effects.

Figure 1.1: Illustration of DEA Estimator

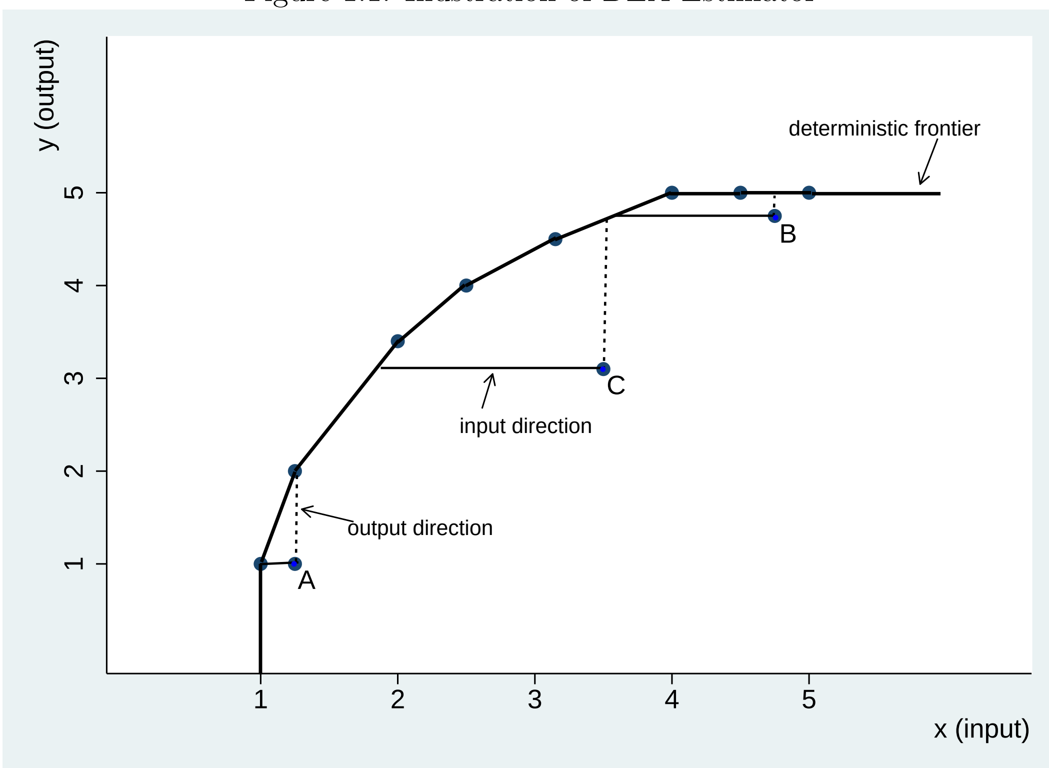
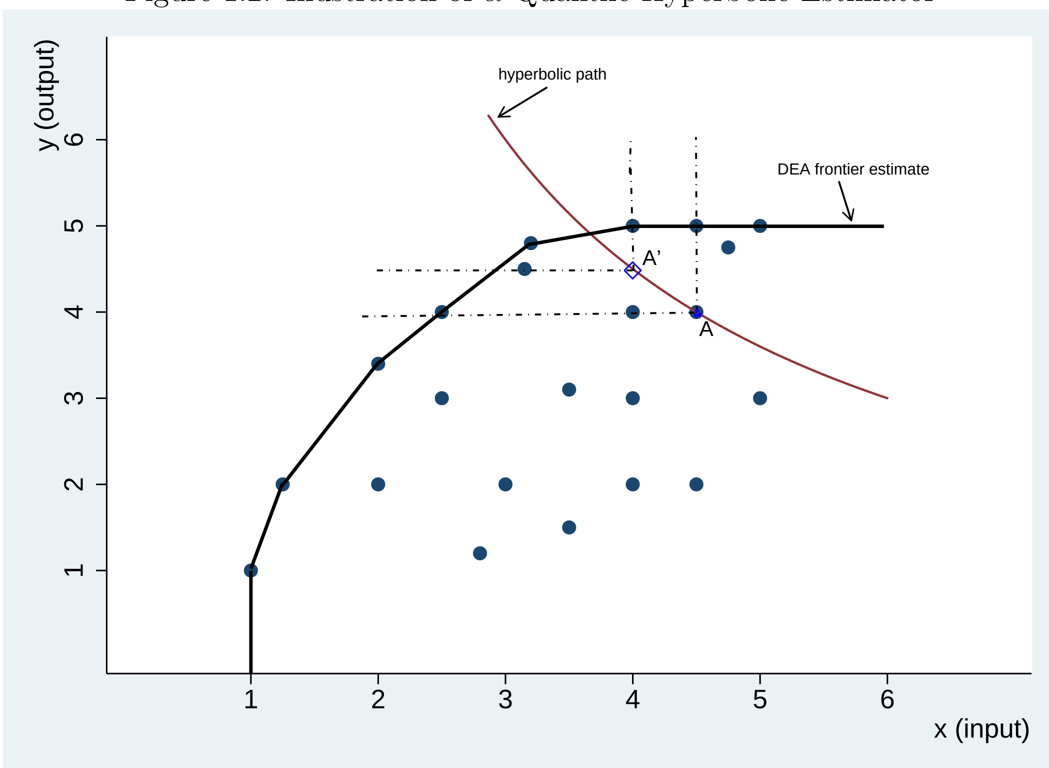


Figure 1.2: Illustration of α-Quantile Hyperbolic Estimator



Chapter 2

Restricting the Size of Banks May Have Costs

2.1 Introduction

The last two decades have seen tremendous change in the U.S. banking sector. Commercial banks have increased in size and consolidated significantly, reducing their numbers by about half since the mid 1980s. Concurrently, bank holding companies (BHCs) have developed steadily during this period; between 1986 and 2005, both their average size measured by total assets and their number almost doubled. The average size of the 1,208 BHCs operating in 1986 was $3.4 billion. By 2005, the average size of BHCs reached almost $7 billion of assets, with 2,276 BHCs operating in the U.S.

BHCs are companies that own or control one or more commercial banks. Most banks in the U.S. are owned by bank holding companies (BHCs). Currently, about 84 percent of commercial banks in the U.S. are part of a BHC structure. More than 75 percent of small banks with assets of less than $100 million are owned by BHCs;

this percentage increases to 100 percent for large banks with more than $10 billion in assets. About 60 percent of minority-owned banks are owned by BHCs.

A typical BHC consists of the parent holding company and one or more subsidiary banks, and perhaps also non-bank subsidiaries. Large BHCs may have hundreds of subsidiary banks and non-banks, although the trend since the 1990s has been toward consolidation; indeed, some BHCs have consolidated all their subsidiary banks into a single bank with interstate branches. From a legal standpoint, organization as a BHC provides advantages in terms of borrowing money, acquiring other banks and non-bank entities, and issuing stock. On the other hand, the downside to organization as a BHC involves additional regulatory scrutiny beyond that incurred by commercial banks.

Prior to 1994, U.S. federal law prohibited commercial banks from operating branches across state lines. BHCs, however, were allowed to own commercial banks in different states, and hence organization as a BHC allowed bank owners to achieve greater geographic diversification subject to various regulatory restrictions. The Riegle-Neal Interstate Banking and Branching Efficiency Act (IBBEA) of 1994 repealed restrictions on interstate branching, allowing BHCs to consolidate their holdings by merging the portfolios of member banks in different states. Subsequently, the Gramm-Leach-Bliley Act (GLBA) of 1999 repealed the Glass-Steagall Act of 1933 which had prohibited banks as well as BHCs from engaging in various non-banking financial activities such as securities underwriting and dealing, insurance agency and underwriting activities, and merchant banking activities. The GLBA permitted BHCs to operate as financial holding companies (FHCs), allowing existing BHCs to acquire full-service securities firms and insurance companies, as well as allowing securities firms and insurance companies to acquire banks (and thereby become a BHC).[1]

[1]For a BHC to be eligible to declare itself an FHC, all of the BHCs depository institution

The rapid growth of BHCs since the 1990s may be due in large part to passage of the IBBEA and GLBA. Regulators and economists have long worried about the efficacy of allowing financial institutions to grow until they become “too big to fail,” and the recent financial crisis that began in 2007 has regenerated the debate (e.g., see Reich, 2008). Various proposals have been discussed in the U.S. Congress that would limit the size and activities of BHCs and FHCs while increasing regulatory scrutiny. This cause has been taken up in the press (e.g., Goodman, 2008; ODriscoll, 2009). Johnson (2009), writing in the New York Times Economix blog, states:

“If you are too big to fail, credit markets see you as lower risk and as a more attractive investmentenabling you to obtain more financing on cheaper terms, and thus become even larger. Everyone agrees, in principle, that this is a bad arrangement. Its an unfair distortion of markets, giving huge banks the opportunity to grow bigger, because they have implicit government guarantees.”

While there may well be good reasons to worry about moral hazard and other issues arising from BHCs that might be regarded as too big to fail, there are also potential costs from limiting the size of BHCs. We examine returns to scale among U.S. BHCs using fully non-parametric regression techniques and find substantial evidence of increasing returns to scale up to the largest BHCs operating today. Our empirical findings suggest that limiting the size of BHCs would result in costs to society in terms of foregone economies, and that these losses would not be trivial. While we do not estimate the benefits of limiting BHCs sizesthis would involve substantial uncer-

subsidiaries must be well-capitalized and well-managed and have satisfactory or better ratings under the Community Reinvestment Act. Small BHCs have a different legal treatment than the other BHCs. Small BHCs are exempt from the consolidated BHC capital guidelines to which larger organizations are subject. The capital adequacy of small BHCs is based on the banks capitalization, just as if the BHC were not present. This means that the BHC, within reasonable parameters determined by its ability to service and retire debt, can use lesser forms of capital or debt funding to provide, for example, equity capital to the bank or to help fund an acquisition. In addition, small BHCs also enjoy simplified reporting requirements. See the Report of Governors (2003) and Heller and Fein (2006) for additional details and discussion.

taintycosts such as those that we find would necessarily reduce any net benefits, and might outweigh any such benefits depending on their magnitude.

It is apparently widely believed that scale economies in banking are exhausted at a relatively small size; Johnson (2009) remarks that "There is no evidence for economies of scale or scopeor other social benefitsfrom banks with assets above $100 billion." Most researchers have found scant evidence of significant scale economies in banking; a number of early studies, using data on U.S. commercial banks, found that scale economies are exhausted at $100-$200 million of total assets. Mester (2005) notes that recent changes in regulation, in particular the changes in branching restrictions, may account for some of the differences in terms of estimated scale economies between earlier and more recent studies. However, many of these studies estimated parametric cost functions that fail the most basic specification tests.[2] Using non-parametric and semi-parametric methods, McAllister and McManus (1993) and Wheelock and Wilson (2001) found that commercial banks face increasing returns to scale up to at least $500 million of total assets. Feng and Serilitis (2009) also find that large commercial banks face increasing returns to scale in a study of 292 commercial banks with at least $1 billion of total assets during 2000-05. Their study relies on Bayesian estimation of a translog output distance function, and avoids the need for input prices which may be subject to considerable measurement error. As the authors acknowledge, however, the translog specification is suitable only for samples composed of relatively homogeneous firms.

The evidence for returns to scale among BHCs is similarly mixed. Stiroh (2000) used four different translog specifications of cost functions for BHCs to examine the performance of BHCs between 1991 and 1997. Stirohs results suggest modest

[2]See McAllister and McManus (1993) and Mester (2005) for discussions of the older literature on bank scale economies.

economies of scale for all years, with the exception of BHCs with assets between \$200 million and \$300 million, for which modest diseconomies are found. The results suggest that the optimal BHC scale increased in the 1990s, and then stabilized. Evanoff (1998) similarly finds only evidence of minor scale economies after estimating a translog shadow cost function for 164 large banks that are part of a BHC. Hughes et al. (2001) find evidence of scale economies among large BHCs while controlling for risk; their estimates are also derived from a translog specification. Vennet (2002) also estimates a translog cost function specification to analyze a sample of 2,375 European financial institutions from 17 countries for 1995 and 1996, and reports finding no significant evidence of either increasing or decreasing scale economies for large universal banks and financial conglomerates.

Much of the research on scale economies in the banking industry has employed fully parametric translog specifications for cost functions. We find that our data trivially reject the translog specification. Rejection of the translog form for BHCs cost function is hardly surprising; McAllister and McManus (1993) and Wheelock and Wilson (2001, 2010) easily reject the translog specification for bank cost functions, and Wheelock and Wilson (2011) find do so similarly while examining credit unions. Consequently, we use fully non-parametric methods to estimate scale economies. We use our estimates and some "back of the envelope" calculations to obtain an idea of the costs, in terms of foregone scale economies, of restricting the size of BHCs.

In the next section, we present our model of BHCs costs and discuss the measures of scale economies that we estimate. Our methods of estimation and inference are discussed briefly in Section 2.3, with technical details given in Appendix A. Section 2.4 presents our discussion of the empirical results, and conclusions are given in the final section.

2.2 The Cost Model

To estimate scale economies we must first specify a model of costs for BHCs. Two issues are involved: (i) the choice of appropriate variables, and (ii) given these variables, the mapping of output quantities, input prices, and other arguments of the cost relationship.

With regard to variable specification, we define the following variables giving quantities of outputs: real estate loans (Y_1), commercial loans (Y_2), consumer loans (Y_3), other loans (Y_4), securities (Y_5), and off-balance sheet items (Y_6) consisting of total non-interest income minus service charges on deposits.[3] We consider three variable input quantities: (i) purchased funds, consisting of the sum of total time deposits of $100,000 or more, allowance for loan and lease losses, and allocated transfer risk reserves less the difference between total assets and the sum of total deposits and total equity capital; (ii) core deposits, consisting of total deposits less time deposits of $100,000 or more; and (iii) labor services, measured by the number of full-time equivalent employees on payroll at the end of each quarter. We measure the prices of purchased funds (W_1), core deposits (W_2), and labor services (W_3) by dividing total expenditure on the given input by its quantity. Variable cost (COST) is the sum of expenditures on these three inputs. Finally, we define two fixed net-put quantities: physical capital, consisting of premises and other fixed assets (Z_1), and financial equity capital (Z_2).

With the exception of labor input (which is measured as full-time equivalent employees) and off-balance sheet output (which is measured in terms of net flow of income), our inputs and outputs are stocks measured by dollar amounts reported

[3] Of the commonly used measures of off-balance sheet output, this definition is the most consistently measurable across banks and over time; Clark and Siems (2002) discuss alternative measures of off-balance sheet activity.

on bank balance sheets, consistent with the widely used "intermediation" model of Sealey and Lindley (1977).

In addition to the variables defined above, we index quarters 1986.Q3 through 2009.Q2 by setting T = 1 for 1986.Q3, T = 2 for 1986.Q4, . . ., T = 92 for 2009.Q2. Since BHCs are observed at regular, quarterly intervals, we treat T an ordered, categorical variable. In addition, we define a binary dummy variable D, equal to one for BHCs that are held by a parent holding company, or zero otherwise.

Our data for individual banks are from the quarterly Bank Holding Company Performance Reports (Call Reports) for U.S. BHCs for 1986.Q3 through 2009.Q2. We omitted BHCs with missing or negative values for any input or output, and converted dollar values to constant year-2000 prices using the GDP deflator. After pooling the data across quarters, 135,635 observations are available for estimation, with from 918 to 2,219 observations in each quarter. Beginning in 2006.Q1, only BHCs with total assets exceeding $500 million filed Call Reports; we include available data from 2006Q1–2009Q3 to aid estimation, but we focus on 1986Q4, 1995Q4, and 2005Q4 in reporting our results in order to facilitate comparisons across time. Table 2.1 reports summary statistics as of the fourth quarters of 1986, 1995, and 2005 for total assets and the variables described above.

The distribution of total assets among BHCs in the U.S. is extremely wide and skewed to the right. Figure 2.1 shows kernel density estimates for (inflation-adjusted) total assets in 1986.Q4, 1995.Q4, and 2005.Q4. The dotted curve gives the density estimate for 1986.Q4; the dashed curve for 1995.Q4, and the solid curve represents 2005.Q4. Total assets are measured in 1,000s of year-2005 U.S. dollars. The densities for each period are skewed to the right, despite the use of a log scale on the figures horizontal axis, although the skewness is less than what one finds for commercial banks in the U.S. over a similar period (see Wheelock and Wilson, 2001, 2010 for comparison). The density estimates also reveal that the distribution of BHC sizes has shifted to the right, reflecting the increase in mean (and median) BHC size over time.

The variables defined above suggest the following mapping:

$$(Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Z_1, Z_2, W_1/W_3, W_2/W_3, T, D) \rightarrow COST/W_3, \tag{2.1}$$

where COST, W_1, and W_2 have been divided by the price of labor services, W_3, to maintain homogeneity with respect to input prices. In order to reduce dimensionality, which has a heavy cost in terms of convergence rates of our non-parametric estimator, we sum the four loan variables; the mapping in (2.1) then suggests the following regression function:

$$\frac{COST}{W_3} = C(\mathbf{y}, \mathbf{w}) + \epsilon, \tag{2.2}$$

where: $\mathbf{y} = [(Y_1 + Y_2 + Y_3 + Y_4)\ Y_5\ Y_6]$, $\mathbf{w} = [\frac{W_1}{W_3}\ \frac{W_2}{W_3}\ Z_1\ Z_2\ T\ D]$, and ϵ is a random error with $E(\epsilon) = 0$ and $VAR(\epsilon)=\sigma^2(\mathbf{y}, \mathbf{w})$. Given that the expectation of ϵ equals 0, $C(\mathbf{y}, \mathbf{w}) = E(COST \mid \mathbf{y}, \mathbf{w})$ is a conditional mean function that can be estimated by various regression techniques.

Now consider a particular point $(\mathbf{y}_0, \mathbf{w}_0)$ in the space $(\mathbf{y}, \mathbf{w})$. The set of points $\mathcal{R}_0=(\theta\mathbf{y}_0, \mathbf{w}_0) \mid \theta \in (0, \infty)$ comprises a ray along which the outputs $(Y_1+Y_2+Y_3+Y_4)$, Y_5, and Y_6 are produced in constant proportion to one another. Ray scale economies can be evaluated by examining how expected cost varies along this ray, providing insight into returns to scale along the ray $\mathcal{R}_0$ is given by

$$\eta(\mathbf{y}, \mathbf{w}) \equiv \frac{\partial \log \mathbf{C}(\theta\mathbf{y}, \mathbf{w})}{\partial \log \theta} = \mathbf{\Sigma}_j \frac{\partial \log \mathbf{C}(\mathbf{y}, \mathbf{w})}{\partial \log \mathbf{y}_j}, \tag{2.3}$$

where j indexes the elements in $\mathbf{y}$. The elasticity in (2.3) is the multi-product analog of marginal cost divided by average cost on the ray $\mathcal{R}_0$ with $\eta(\mathbf{y}, \mathbf{w})(<, =, >)1$ implying (increasing, constant, decreasing) returns to scale as outputs in $\mathbf{y}$ are expanded along the ray $\mathcal{R}_0$. Banks for which $\eta(\mathbf{y}, \mathbf{w}) \neq 1$ are not competitively viable; if BHCs are subject to the normal rules of competitive behavior, either a smaller or a larger firm could drive a BHC with $\eta(\mathbf{y}, \mathbf{w}) \neq 1$ from a competitive market.

The measure defined in (2.3) requires estimation of derivatives of the cost function. We employ fully nonparametric estimation methods, as discussed below in Section 2.3. Nonparametric estimates of derivatives of a function are typically noisier than estimates of the function itself.[4] Hence, we define the ratio

$$S(\theta \mid \mathbf{y}_0, \mathbf{w}_0) \equiv \frac{C(\theta\mathbf{y}_0, \mathbf{w}_0)}{\theta C(\mathbf{y}_0, \mathbf{w}_0)}). \tag{2.4}$$

It is straightforward to show that

$$\frac{\partial S(\theta \mid \mathbf{y}_0, \mathbf{w}_0)}{\partial\theta} <=> 0 \Longleftrightarrow \eta(\mathbf{y}_0, \mathbf{w}_0) <=> 1; \tag{2.5}$$

i.e., $S(\theta \mid \mathbf{y}_0, \mathbf{w}_0)$ is decreasing (constant, increasing) in θ if returns to scale are increasing (constant, decreasing) at $(\theta\mathbf{y}_0, \mathbf{w}_0)$ along the ray $\mathcal{R}_0$ passing through $(\mathbf{y}_0, \mathbf{w}_0)$. In addition, $S(1 \mid \mathbf{y}_0, \mathbf{w}_0)= 1$ by definition. Thus, we investigate ray scale economies (RSE) along a ray $\mathcal{R}_0$ by estimating $C(\mathbf{y}_0, \mathbf{w}_0)$ and $C(\theta\mathbf{y}_0, \mathbf{w}_0)$ for various values of θ, and using confidence bands to determine whether $S(\theta \mid \mathbf{y}_0, \mathbf{w}_0)$ is downward or upward sloping.

In the empirical analysis below, we define the fixed point $(\mathbf{y}_0, \mathbf{w}_0)$ by taking medians of the variables in our model. Of course, few if any BHCs may be located along the ray $\mathcal{R}_0$. Although RSE is a convenient measure of scale economies, it may

[4]This is particularly true for the present case where we would require derivatives in several dimensions; in addition, bandwidth selection becomes problematic when estimating derivatives in more than one dimension.

be misleading if most BHCs are located "far" from $\mathcal{R}_0$. As an alternative to RSE, we also consider scale economies along each BHCs expansion path, i.e., along the path which holds each BHCs output mix constant. Consider a BHC operating at the point $(\mathbf{y}_0, \mathbf{w}_0)$, with cost $C(\mathbf{y}_0, \mathbf{w}_0)$. Let γ be a small positive number, say 0.05, and consider how cost changes as we move from $((1-\gamma)\mathbf{y}_0, \mathbf{w}_0)$ to $((1+\gamma)\mathbf{y}_0, \mathbf{w}_0)$; along this path, the output mix remains constant in the sense that relative proportions are maintained. Now let $\theta(1-\gamma)\mathbf{y}_0 = (1+\gamma)\mathbf{y}_0$; then $\theta = (1+\gamma)/(1-\gamma) \approx 1.1053$.

The following expression provides a measure of expansion-path scale economies (EPSE) for a BHC operating at $(\mathbf{y}_0, \mathbf{w}_0)$:

$$\varepsilon_0 = \frac{C(\theta(1-\gamma)\mathbf{y}_0, \mathbf{w}_0)}{\theta C((1-\gamma)\mathbf{y}_0, \mathbf{w}_0)} = \frac{C((1+\gamma)\mathbf{y}_0, \mathbf{w}_0)}{(\frac{1+\gamma}{1-\gamma})C((1-\gamma)\mathbf{y}_0, \mathbf{w}_0)}. \tag{2.6}$$

A BHC operating at $(\mathbf{y}_0, \mathbf{w}_0)$ experiences (decreasing, constant, increasing) returns to scale along the path from $((1-\gamma)\mathbf{y}_0, \mathbf{w}_0)$ to $((1+\gamma)\mathbf{y}_0, \mathbf{w}_0)$ as $\varepsilon_0(>, =, <)1$. Our measure ε_0 provides an indication of returns to scale faced by a particular BHC along the path from the origin through the BHCs *observed* output vector, starting at a level equal to 95-percent of the quantities in $\mathbf{y}_0$ and continuing to a level equal to 105-percent of the quantities in $\mathbf{y}_0$, or 110.53 percent of the starting point at $(1-\gamma)\mathbf{y}_0$ for γ= 0.05. Hence ε_0 measures the increase in cost resulting from an increase in output by a factor of θ; when output increases by a factor of θ, cost increases by a factor (ε_0 x θ).

The RSE and EPSE measures are both defined in terms of a BHCs cost function. The following section discusses a strategy for estimating the cost function non-parametrically, which in turn allows us to estimate, and make inference about, these measures of scale economies.

2.3 Estimation and Inference

Various approaches exist for estimating regression functions (i.e., conditional mean functions) such as the one defined above in (2.2). A common approach, particularly in banking studies, is to estimate the conditional mean function parametrically using a translog specification. However, because the translog function is merely a quadratic specification in log-space, this approach limits the variety of shapes the cost function is permitted to take. Further, because the translog is derived from a Taylor expansion of the cost function around the mean of the data, it makes little sense to use a translog specification to attempt inference about returns to scale over units of widely varying size.

In order to test our suspicions about the translog specification, we performed separate specification tests using data for each of the 92 periods (1986Q3–2009Q2) represented in our data. For each period, we divided the data into two subsets consisting of BHCs with total assets less than or equal to median total assets for the given period, and BHCs with total assets greater than the median. Using the variables listed in (2.1), but not summing the four loan variables Y_1, Y_2, Y_3, and Y_4 but rather including them as separate variables, for each period we estimated separate translog cost functions using the two sub-samples, and then performed a Wald test using Whites (1980) heteroskedasticity-consistent covariance matrix estimator to test equivalence of parameters vectors across the two sub-samples. We rejected the null hypothesis of equality in each of the 92 cases, with p-values ranging from $10^{11.51}$ to $10^{-211.60}$.

Rejection of the translog functional form is hardly surprising. Several studies have noted that the parameters of a translog function are unlikely to be stable when the function is fit globally across units of widely varying size.[5] The problem suggests the use of nonparametric estimation methods. Although nonparametric methods are less efficient than parametric methods in a statistical sense when the true functional form is known, nonparametric estimation avoids the risk of specification error when the true functional form is unknown, as in the present application.

Given that the translog specification is trivially rejected by our data, we use a fully nonparametric, local-linear estimator augmented to handle the discrete covariate D and the ordered categorical variable T along the lines used by Wheelock and Wilson (2011). Nonparametric regression models may be viewed as infinitely parameterized; as such, any parametric regression model (such as the translog cost function) is nested within a nonparametric regression model. Clearly, adding more parameters to a parametric model affords greater flexibility. Nonparametric regression models represent the limiting outcome of adding additional parameters, and can be viewed as nesting various parametric models such as the translog specification as well as others.[6]

[5]See, for example, Guilkey et al. (1983) and Chalfant and Gallant (1985) for Monte Carlo evidence, and Cooper and McLaren (1996) and Banks et al. (1997) for empirical evidence involving consumer demand. Still others have found a similar problem while estimating cost functions for hospitals (Wilson and Carey, 2004), for US commercial banks ((McAllister and McManus, 1993; Mitchell and Onvural, 1996; and Wheelock and Wilson, 2001), and for credit unions (Wheelock and Wilson, 2010); hospitals, banks, and credit unions all vary widely in terms of size.

[6]Fan and Gijbels (1996, chapter 1) and Hardle and Linton (1999) give nice descriptions of nonparametric regression and the surrounding issues. Note that several possibilities for nonparametric regression exist. For example, orthogonal series estimators based on the ideas of Szeg (1959) and Gallant (1981, 1982) involve representing the conditional mean function by an infinite Fourier series and using orthogonal polynomials (e.g., Laguerre or Legendre polynomials) or other functions (e.g, transcendental functions or Muntz-Satz expansions) to represent the basis functions, and have been used in studies of bank costs and elsewhere. One must choose a truncation point for the Fourier series; cross-validation and other methods (e.g., Eastwood, 1991) may be used, although published papers using this approach to analyze banks have typically not optimized the number of terms according to such criteria. In addition, Barnett et al. (1991) note that the basis functions with which Gallants model seeks to span the neoclassical function space are sines and cosines, despite the fact that such trigonometric functions are periodic and hence are far from neoclassical. In other words, the basis functions, which should be dense in the space to be spanned do not themselves even lie within that space. Instead of trigonometric functions, one could use as the basis functions members of a family of orthogonal polynomials, but the problems of determining the optimal number of terms,

Precise details of our estimation procedure are given in the Appendix. We use a dimension-reduction technique along the lines of Wilson and Carey (2004) and Wheelock and Wilson (2011) to help mitigate the slow convergence rate of our estimator; the technique involves transforming the data to principal-components space and then dropping the principal components that are nearly redundant in terms of the independent linear information that they contain. We incorporate the binary dummy variable D and the ordered categorical variable T into our estimation by augmenting the kernel weights in our local linear estimator with additional kernel functions for these variables as used by Racine and Li (2004), Wilson and Carey (2004), and Wheelock and Wilson (2011). This approach requires three bandwidth parameters, with one controlling the degree of smoothing among the continuous covariates, one controlling the degree of smoothing across the two sub-samples defined by the binary variable D, and with the third controlling the degree of smoothing over time, represented by T. Inference is made using the wild bootstrap proposed by Hardle (1990) and Hardle and Mammen (1993). See the Appendix for further details.

2.4 Empirical Results

As discussed in the Appendix, values for the three bandwidth parameters h_0, h_1, and h_2 are needed for estimation. Using the Nelder and Mead (1965) simplex algorithm with the BHC data to optimize the least-squares cross-validation function in (A.15) yields $h_0 = 0.00768236$, $h_1 = 0.974708$, and $h_2 = 0.657994$. Given our sample size of $n = 135{,}635$ and our use of a nearest-neighbor bandwidth for the continuous dimensions in our model, we have $k = [h_0 n] = 1{,}041$ so that smoothing is over the 1,041

and using these in a non-linear, maximum-likelihood framework, remain. As a practical matter, in the present multivariate setting with a large number of observations, this method would incur the numerically challenging problem of inverting very large moment matrices. Our local-linear estimator avoids these problems.

nearest neighbors of a given, fixed point. As discussed in the Appendix, smoothing in the continuous dimensions is accomplished using a spherical Epanechnikov kernel function; consequently, observations closer to the point of interest receive more weight, while those farther way receive less weight. Only the 1,041 observations closest to the given point of interest receive non-zero weight in the estimation.

Table 2.2 presents summary statistics for our estimates of EPSE by size quartile for each of the three periods we consider. Recalling that values of the EPSE measure defined in (2.6) less than one indicate increasing returns to scale, the results suggest evidence of increasing returns. The results in Table 2.2, however, merely describe the empirical distributions of point estimates. In order to test statistical significance, we used the bootstrap procedure described in the Appendix to estimate 95-percent confidence intervals for ε_i in each of the three periods. We then tabulated, for each size-quartile in each period, how many observations had confidence intervals lying to the left of one, covering one, or lying to the right or one.

The results of this analysis are shown in Table 2.3, where the totals are tallied for each size–quartile in each period. In addition, the confidence interval estimates are plotted in Figures 2.4-2.6 after sorting the interval estimates by their upper bounds in each size quartile. The results are striking. In each period, almost all BHCs are shown to face significant increasing returns to scale, in all size groups. Constant returns to scale cannot be rejected for only a few BHCs, and there is not evidence that *any* BHCs face decreasing returns to scale. Moreover, for those that face (statistically) significant increasing returns to scale, Figures 2.4-2.6 reveal that the estimated confidence intervals upper bounds like well below one (i.e., the distance between one and the upper bound of each confidence interval estimate is typically much greater than the width of the estimated confidence intervals.

In 2005.Q4, for only 15 BHCs in the largest size-quartile can constant returns to scale *not* be rejected. One might reasonably ask how many BHCs in the same size-quartile are larger than each of these 15 BHCs. Among these 15 BHCs, the number of BHCs that are larger ranges from 3.52 percent to 97.71 percent of the BHCs in the same size-quartile; for 6 of the 15 BHCs, more than half of BHCs in the same size-quartile are larger. The 15 BHCs in the largest size-quartile for which constant returns cannot be rejected are not the largest BHCs. The evidence for increasing returns to scale among the very largest of BHCs seems quite strong.

We performed a similar analysis on BHCs for the last two quarters covered by our data (i.e., 2009.Q1 and 2009.Q2), although in Tables 2.2-2.3 the last period we report for is 2005.Q4 since after 2005.Q4, only banks with more than $500 million of total assets reported FR-Y9C data as discussed earlier. The results for 2009.Q1 and 2009.Q2, however, are very similar to those in the other quarters we examined. Table 2.4 shows total assets (in current dollars) of the 12 largest BHCs operating in the U.S. as of June 30, 2010; all of these except the eighth largest, Barclays Group US Inc., appear in our sample (Barclays is missing due to missing data needed to construct our variables). In 2009.Q1, the largest BHC for which we cannot reject constant returns to scale in favor of increasing returns using our EPSE estimates and corresponding estimated 95-percent confidence intervals is Taunus Corporation, with total assets of $372.75 billion (2010 dollars) at the end of 2009.Q1. The next-largest BHC for which we cannot reject constant returns is South Plains Financial, Inc., with total assets of $1.98 billion (2010 dollars) at the end of 2009.Q1. For 2009.Q2, South Plains Financial, Inc. is the largest BHC for which we cannot reject constant returns to scale. Consequently, we find evidence of increasing returns up to and including the largest BHCs operating in the U.S.

Our estimates of RSE are illustrated in Figure 2.2 for $D = 0$ (i.e., for BHCs that are not held by another holding company) and in Figure 3 for $D = 1$ (i.e., for

BHCs that are held by another holding company. In both figures, RSE estimates are plotted as a solid line, while bounds of corresponding 95 percent confidence interval estimates are plotted with dashed lines. On the horizontal axes, θ (plotted on a logarithmic scale) serves as a scaling parameter, with θ =1 corresponding to the "median BHC" in each period.

The results in Figures 2.2-2.3 are very similar across the two groups of BHCs. Recalling that the RSE measure is defined in (2.4) so that downward sloping relationships indicate increasing returns to scale along the ray from the origin through the medians of the data, it is apparent that the results indicate increasing returns to scale throughout the range of BHC sizes in each period, Hence our results for RSE are consistent with our results for EPSE. Overall, we find strong evidence that increasing returns to scale prevail throughout the size distribution of BHCs.

As discussed in Section 2.1, recent policy discussions among regulators and lawmakers have considered the idea of restricting the size of large BHCs. While there may be good reasons for doing so, our evidence points to some of the costs that would result from such policies. Table 2.5 shows our EPSE estimates and corresponding 95-percent confidence interval estimates for the four BHCs with total assets exceeding \$1 trillion in 2009.Q1 and 2009.Q2. The estimates are similar across the two quarters as well as across the four BHCs. The average of the EPSE estimates shown in Table 2.5 is 0.9162; the averages of the lower and upper bounds of the estimated 95-percent confidence intervals are 0.9082 and 0.9276, respectively.

Recall from the discussion in Section 2.2 following (2.6) that the expansion-path scale economy measure measures the increase in costs when output increases by a factor θ. If output increases by a factor θ, then costs increase by a factor (εxθ). Alternatively, if output decreases by a factor $1/\theta$, then costs decrease by a factor $1/(\varepsilon \mathrm{x} \theta)$; since $\theta \approx 1.1053$, a decrease in output levels by a factor $1/\theta \approx 0.9047$ leads to a reduction in costs by a factor (εx1.1053). For Bank of America Corp. to reach a size of \$1 trillion of assets, it would have to shrink by a factor of about $1/2.366 \approx 0.4227$. Setting $0.9047^a = 0.4227$ implies a ≈ 8.5979; using the average of the EPSE estimates reported in Table 5, we could expect Bank of America Corp.s variable costs to shrink by a factor of about $(0.9162\text{x}1.1053)^{8.5979} \approx 0.8974$. In other words, shrinking Bank of America Corp to a size of \$1 trillion of total assets would likely reduce its quarterly variable cost by only about 89.74 percent, or from \$15.456 billion (i.e., the average of its quarterly costs incurred in 2009.Q1 and 2009.Q2) to about \$13.870 billion–a reduction of about \$1.586 billion.

Similar calculations using the average of the EPSE estimates shown in Table 2.5 reveal that shrinking the other three BHCs listed in Table 2.5 (i.e., J.P. Morgan Chase, Citigroup, and Wells Fargo) to \$1 trillion of total assets would reduce their quarterly variable costs by 91.58, 92.03, and 97.47 percent (respectively). Using averages of costs for the first two quarters of 2009 shown in Table 2.5, costs would be reduced from about \$10.447 billion to about \$9.567 billion for J.P. Morgan Chase; from about \$10.293 billion to \$9.473 billion for Citigroup; and from about \$8.462 billion to \$8.248 billion for Wells Fargo. Overall, shrinking each of the four largest BHCs to \$1 trillion of assets would reduce costs by a total of about \$3.500 billion *per quarter*. However, the values shown in Table 2.4 show that shrinking the four largest BHCs to \$1 trillion of total assets each would leave \$3.544 trillion of assets to be held by other BHCs. To give a conservative estimate (in view of our empirical evidence of increasing returns to scale throughout the size-distribution of BHCs), suppose that these assets are placed in 3.544 BHCs with assets of \$1 trillion. The average of the estimated quarterly variable costs after shrinking each of the four largest BHCs as described above is \$10.290 billion; 3.544 times this amount is \$36.466 billion per quarter. Subtracting the amount saved in each of the four largest existing

BHCs leaves about \$32.966 billion per quarter, or about \$131.864 billion per year. While there may be good reasons to consider limiting the size of BHCs, our results indicate that there would be non-trivial costs resulting from such a policy in terms of foregone opportunities to exploit returns to scale, resulting in a mis–allocation of resources, *ceteris parabus.*

2.5 Conclusion

We find strong evidence of increasing returns to scale throughout the size range of BHCs operating in the U.S. We have used a fully non-parametric estimation procedure, after demonstrating that the widely used translog cost function specification is easily rejected by our data. Use of fully non-parametric methods avoids any risk of specification error, but comes at a cost of increased uncertainty surrounding our estimates. Nonetheless, we find statistically significant evidence of increasing returns to scale among the very largest BHCs.

While considering our rough estimate in Section 2.3 of constraining BHCs to no more than \$1 trillion of assetsabout \$131.864 billion per yearone should remember that financial crises happen not every year, but somewhat infrequently. These costs (in terms of foregone scale economies) would accumulate to \$1.329 trillion after ten years, and to \$2.637 trillion after twenty years. The cost of the recent financial crisis is difficult to quantify, but our results suggest that the cost of constraining the size of BHCs is quite large over time. Moreover, the benefits of constraining the size of BHCs is far from certain in terms of how, whether, and to what degree such policy might reduce the risk of future financial crises. In this paper, we have provided only part of the analysis that should be considered in ongoing policy discussions, but we believe it is an important piece of the puzzle.

Table 2.1: Summary statistics

	Min	1st Quartile	Median	3rd Quartile	Max
			–1986.Q4–		
COST	515.39	4973.15	9189.47	31338.16	6852700.91
Y_1	517.23	35905.08	87393.46	269712.62	67975240.00
Y_2	188.23	22886.18	58442.10	208152.12	64391020.00
Y_3	26.89	16327.11	39410.61	137146.67	54984416.00
Y_4	0.00	6123.70	17680.68	71324.57	27922442.00
Y_5	3136.59	52261.45	112549.44	331323.35	28526666.00
Y_6	5806.56	120.21	422.32	2152.35	1738331.38
Z_1	53.78	3279.34	7492.70	22195.36	5025185.50
Z_2	115546.05	11373.50	27057.23	88256.70	14463423.00
W_1/W_3	157.83	419.92	497.05	627.26	13827.08
W_2/W_3	0.01	1.12	1.39	1.70	70.29
ASSETS	19626.22	192941.34	374067.08	1240296.22	310217024.00
			–1995.Q4–		
COST	393.55	4051.10	6634.98	19438.08	5697120.13
Y_1	0.00	88200.17	153113.57	423406.27	82565792.00
Y_2	0.00	21198.95	42328.79	124642.59	58636196.00
Y_3	0.00	14962.73	33318.19	105555.92	66081012.00
Y_4	0.00	2434.61	9533.55	40925.40	24542454.00
Y_5	6116.35	85182.36	136944.40	367828.86	93733264.00
Y_6	1686.89	195.72	514.38	2015.64	2538284.00
Z_1	124.77	4495.82	8054.62	20717.29	5390948.50
Z_2	3403.14	24035.10	38973.36	122362.11	24736954.00
W_1/W_3	335.82	863.56	1032.20	1336.49	278067.07
W_2/W_3	0.11	1.38	1.62	1.95	304.54
ASSETS	32560.98	259203.45	423738.08	1245432.28	314200416.00

Table 2.1 – continued

	Min	1st Quartile	Median	3rd Quartile	Max
			–2005.Q4–		
COST	446.43	2165.05	3425.43	7112.91	14311373
Y_1	0	111115.1	190509.4	385150.4	324374976
Y_2	0	18065.51	35103.69	73747	126887928
Y_3	0	5146.81	10821.16	25205.49	172263904
Y_4	0	1296.31	6235.86	20565.65	104254976
Y_5	0	45128.07	78982.3	163520	383155776
Y_6	423391.12	163.62	368.89	1072.7	9895471
Z_1	64.46	3980.9	7148.66	14412.59	9004988
Z_2	3260.48	19978.86	31849.16	64492.36	119586840
W_1/W_3	734.92	2368.26	2982.66	3925.15	2241650
W_2/W_3	0.02	1.56	1.88	2.31	1109.73
ASSETS	27413.6	231925.6	367558.4	735852.4	1481531264

NOTE: All variables (except binary dummy variable D) are measured in 1,000s of U.S. year-2005 dollars.

Table 2.2: Summary Statistics for Expansion-Path Scale Economy Estimates by Size-Quartile

Size Quartile	Min	1st Quartile	Median	Mean	3rd Quartile	Max
			–1986.Q4–			
1	0.9211	0.9351	0.9369	0.937	0.9387	0.9579
2	0.8893	0.9341	0.9356	0.9357	0.9371	0.977
3	0.8419	0.9322	0.934	0.9333	0.9358	0.9606
4	0.8959	0.926	0.9288	0.928	0.9311	0.9442
			–1995.Q4–			
1	0.8782	0.9265	0.9315	0.9316	0.936	1.0217
2	0.8851	0.9319	0.9345	0.9344	0.9378	0.9595
3	0.8932	0.9302	0.9335	0.9332	0.936	0.9723
4	0.8207	0.9234	0.927	0.9266	0.9302	0.9527
			–2005.Q4–			
1	0.8595	0.9131	0.9224	0.9221	0.9312	0.9684
2	0.8675	0.9264	0.9334	0.9324	0.94	0.9663
3	0.8848	0.9274	0.9326	0.9329	0.9385	0.9697
4	0.8819	0.923	0.9275	0.9285	0.9335	0.9627

NOTE: For each period, summary statistics are given for the first, second, third, and fourth quartiles of BHCs total assets.

Table 2.3: Expansion–Path Scale Economies, 95–percent Significance

Quartile	IRS	CRS	DRS
	–1986–		
1	277	1	0
2	271	6	0
3	271	6	0
4	275	3	0
	–1995–		
1	319	1	0
2	316	3	0
3	318	1	0
4	316	3	0
	–2005–		
1	553	15	0
2	554	14	0
3	554	14	0
4	553	15	0

Table 2.4: 12 Largest Bank Holding Companies as of 30 June 2010

BHC	Total Assets
Bank of America Corp.	$2,366,086,945,000
J.P. Morgan Chase & Co.	2,014,019,000,000
Citgroup, Inc.	1,937,656,000,000
Wells Fargo & Co.	1,225,862,000,000
Goldman Sachs Group, Inc.	883,529,000,000
Morgan Stanley	809,456,000,000
Metlife, Inc.	573,907,057,000
Barclays Group US, Inc.	356,186,000,000
Taunus Corporation	348,586,000,000
HSBC North America Holdings, Inc.	333,997,956,000
U.S. Bancorp	283,243,000,000

NOTE: Total assets measure in current (2010) dollars. Data were obtained from the Federal Reserve Systems National Information Center.

Table 2.5: EPSE Estimates for BHCs with Total Assets Exceeding $1 Trillion

BHC	EPSE	–95% CI–		COST
		–2009.Q1–		
Bank of America Corp.	0.9155	0.9081	0.9282	$16.806
J.P. Morgan Chase & Co.	0.9144	0.9074	0.9246	10.883
Citgroup, Inc.	0.9168	0.911	0.9269	10.847
Wells Fargo & Co.	0.9173	0.9121	0.9262	8.509
		–2009.Q2–		
Bank of America Corp.	0.9164	0.9083	0.9294	14.106
J.P. Morgan Chase & Co.	0.915	0.902	0.9268	10.011
Citgroup, Inc.	0.9164	0.9052	0.9311	9.739
Wells Fargo & Co.	0.9175	0.9111	0.9278	8.415

NOTE: COST is (quarterly) total variable cost, measured in billions of 2010 dollars.

Figure 2.1: Density of (Log) Total Assets for 1986.Q4, 1995.Q4, and 2005.Q4

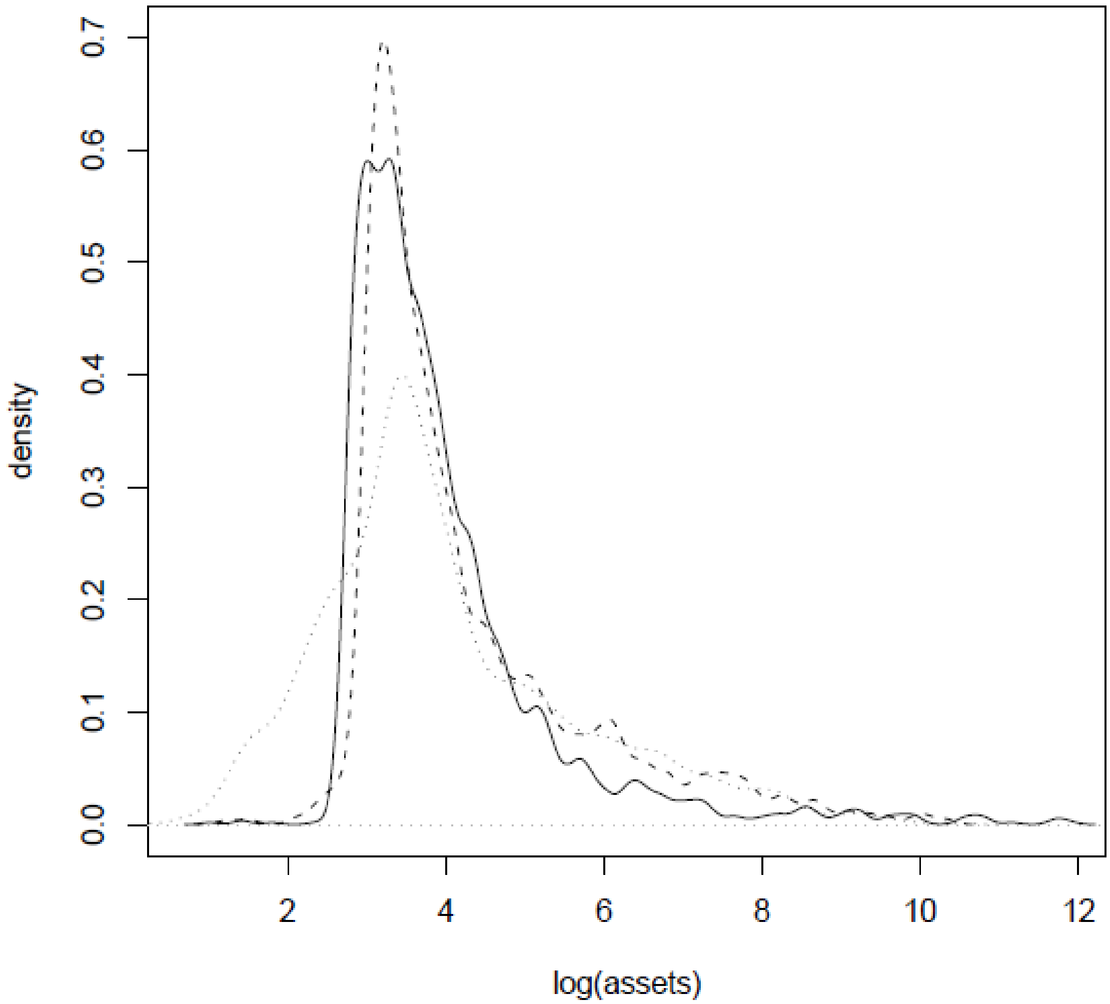


Figure 2.2: Ray Scale Economies (D = 0)

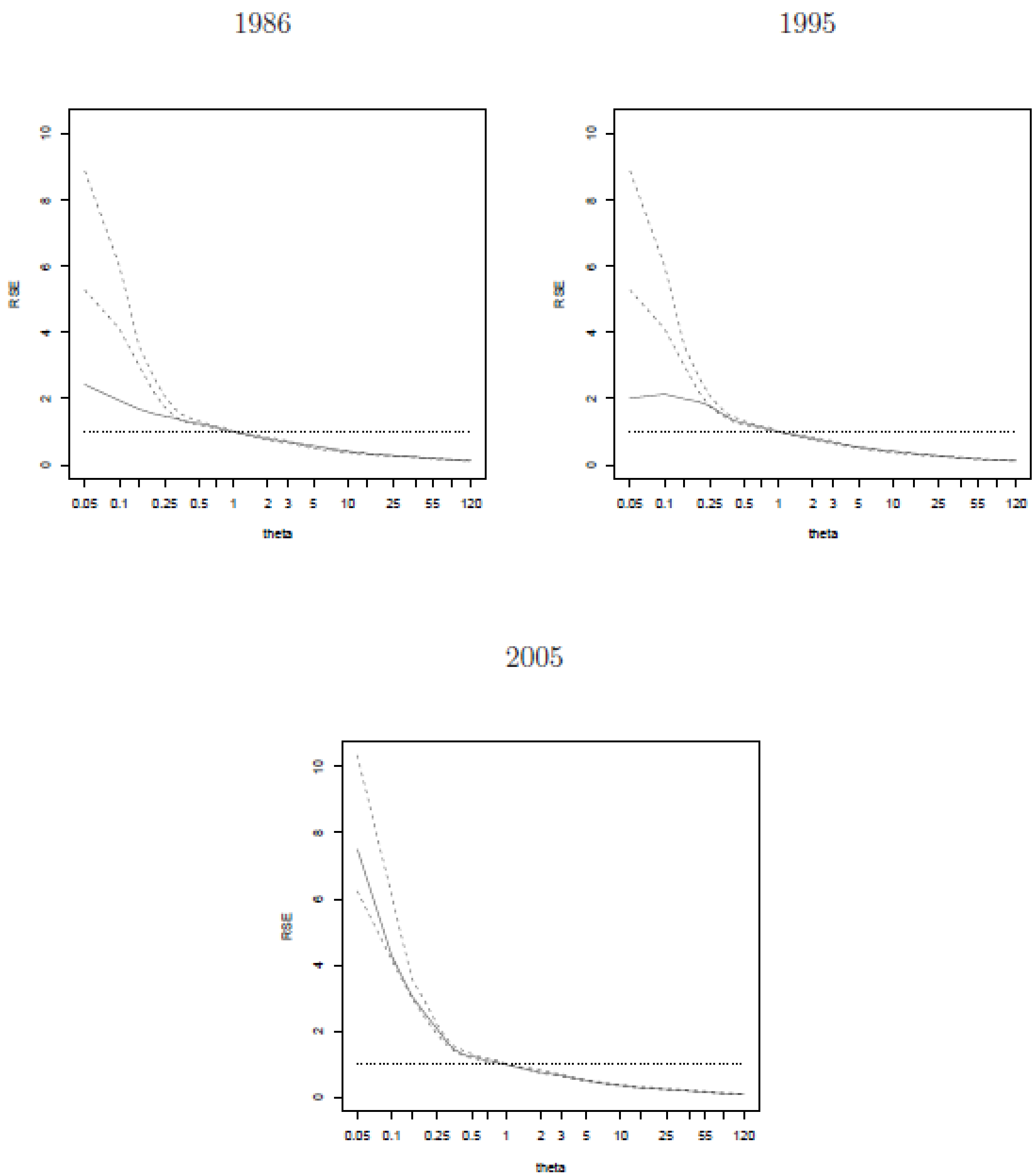


Figure 2.3: Ray Scale Economies (D = 1)

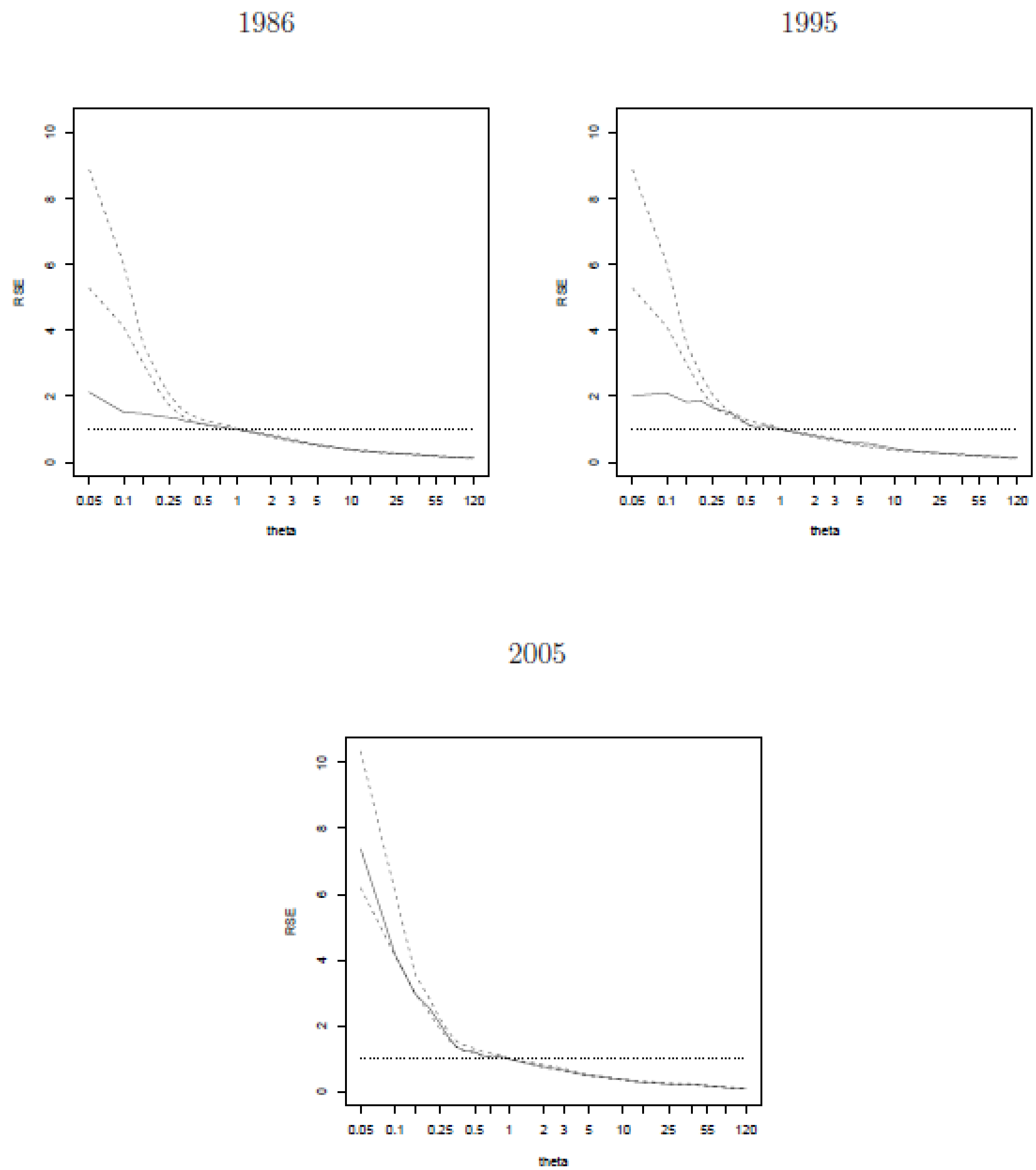


Figure 2.4: Expansion Path Scale Economies by Size–Quartile, 1986

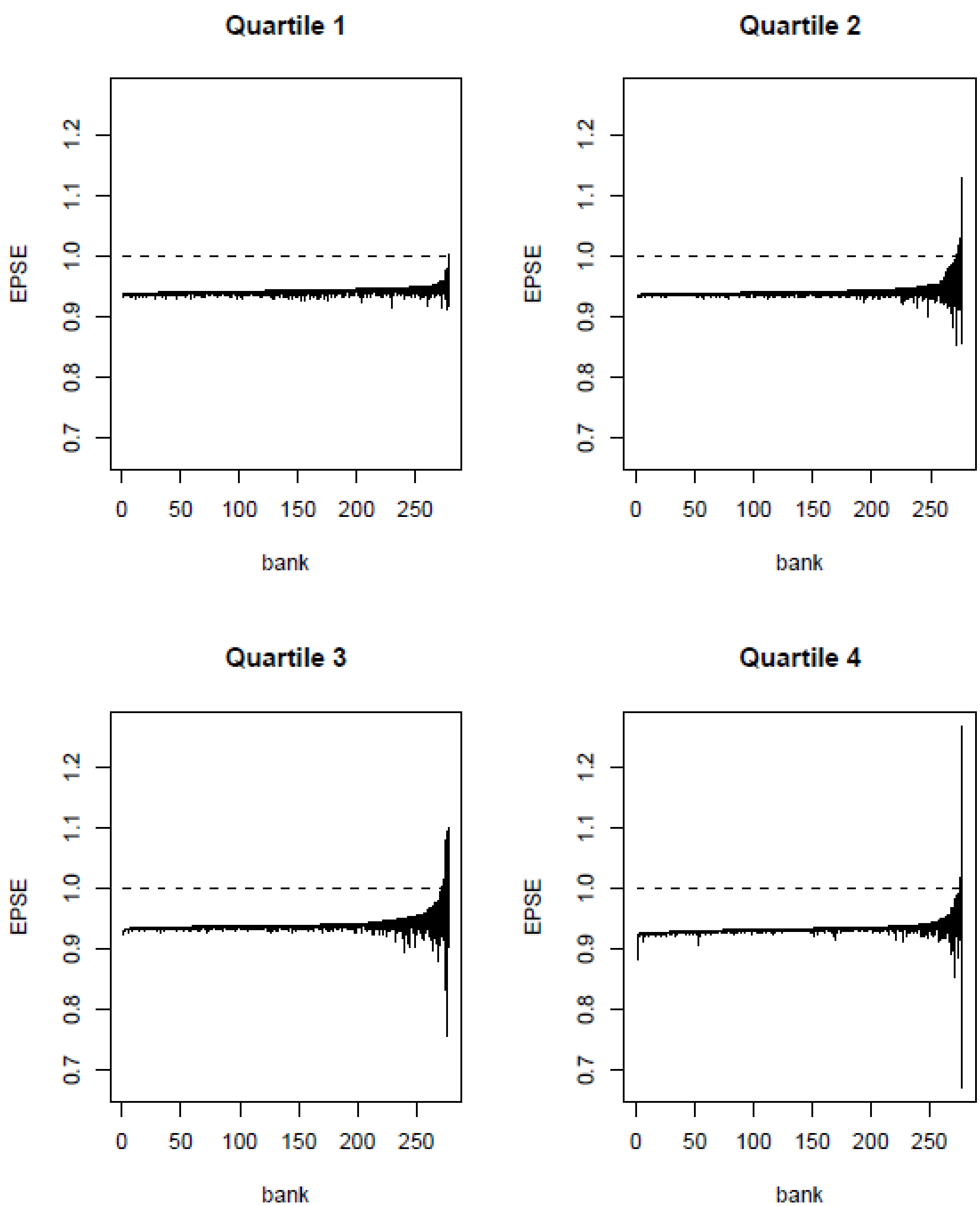


Figure 2.5: Expansion Path Scale Economies by Size–Quartile, 1995

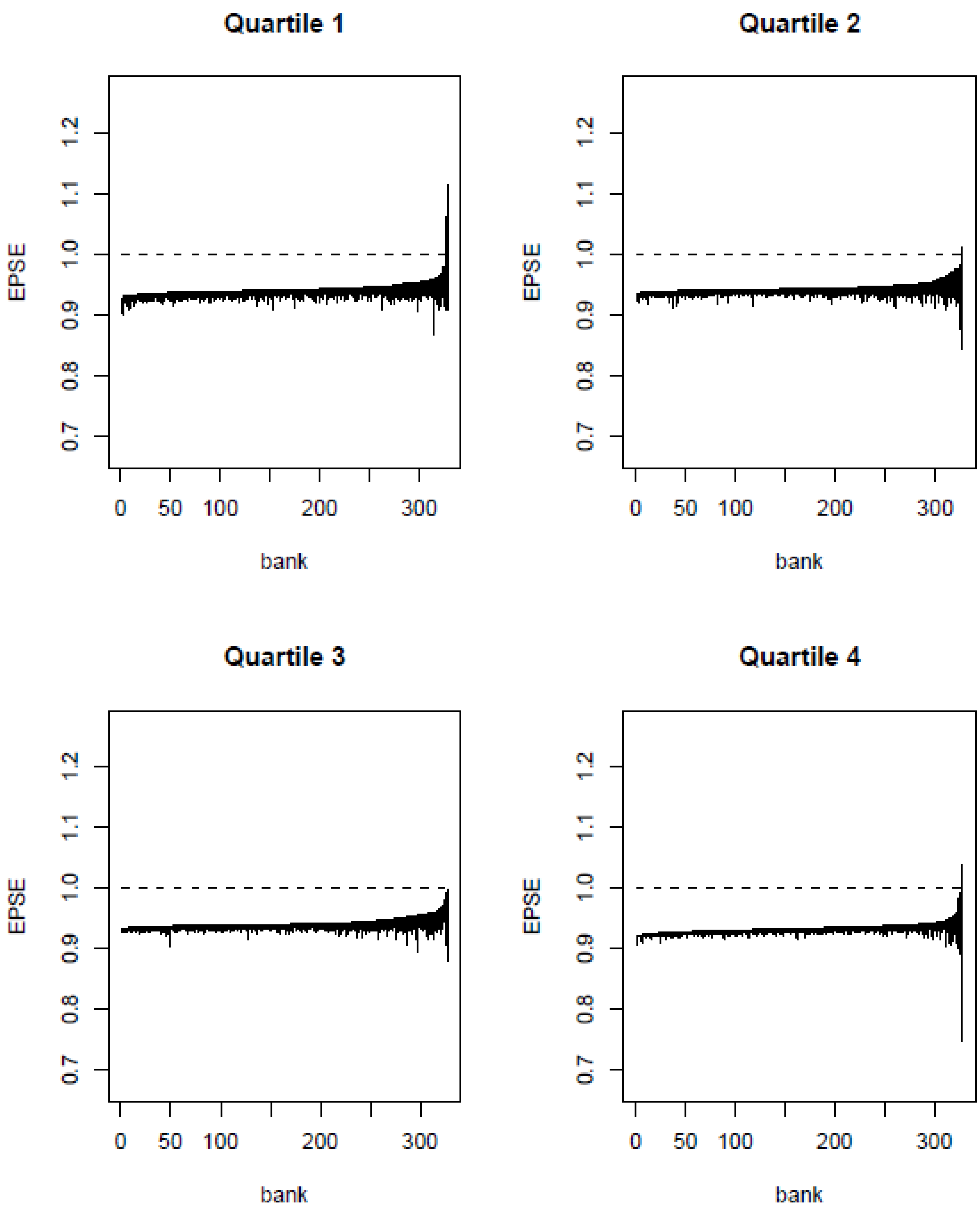
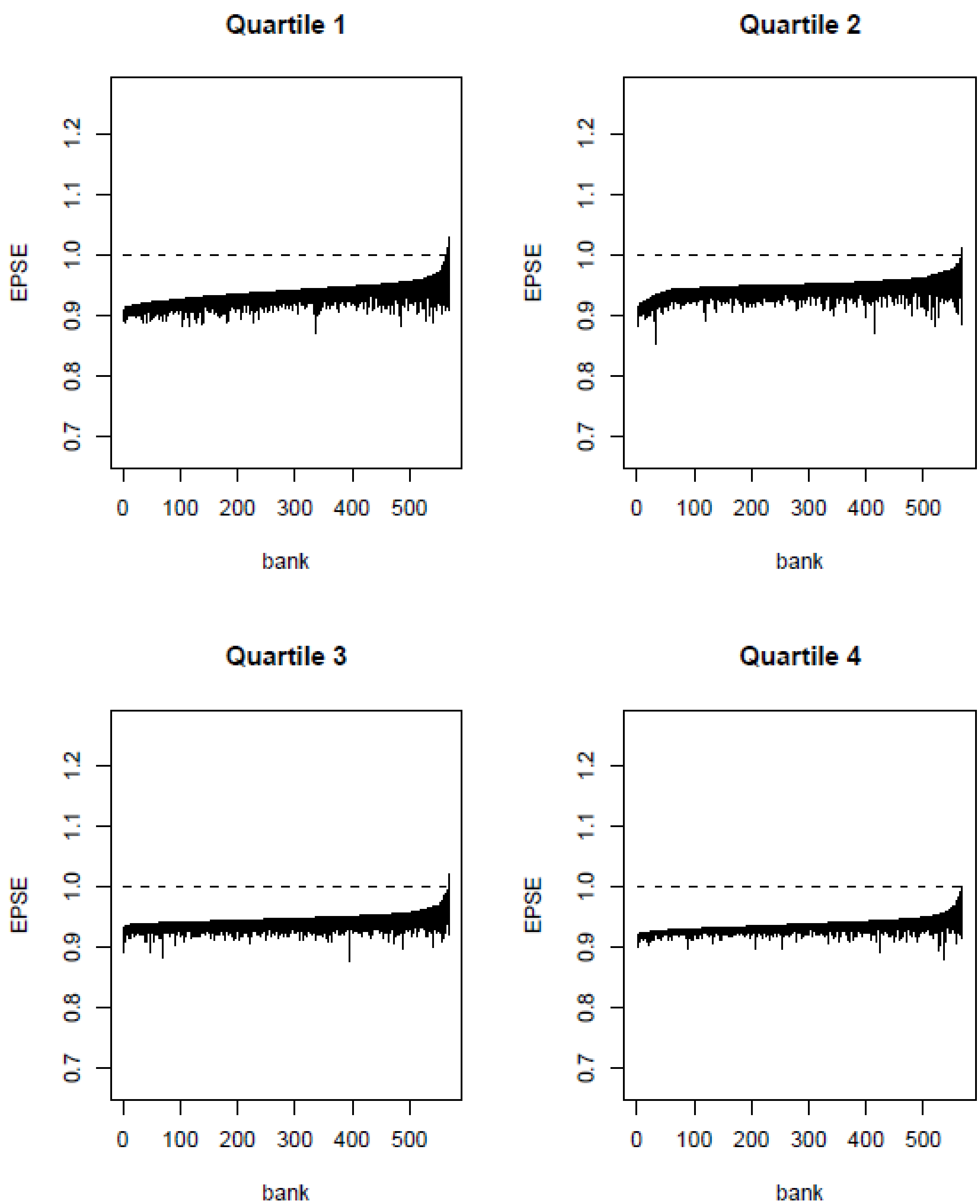


Figure 2.6: Expansion Path Scale Economies by Size–Quartile, 2005



Chapter 3

Evolution of the U.S. Bank Holding Companies' Performance over Time: Evidence from Nonparametric Efficiencies

3.1 Introduction

During the last three decades, the U.S. banking industry has undergone substantial transformations. Major regulatory changes redefined the banking business environment: deregulation of deposits accounts, changes in capital and reserve requirements, liberalization of the intrastate and interstate banking, legal permission to banks to act in the insurance market, securitization of traditional banking assets. In addition, significant technological changes and developments in the financial services sector took place. The progress in telecommunications led to the globalization of the financial industry, and the new technologies enabled innovation and engineering of new products, and improved risk-management techniques. As a result, the banking sector became more competitive, facing strong competition from foreign banks too.

Boyd and Gertler (1993), Berger et al. (1995), Jones and Critchfield (2005) document the transformation of the U.S. commercial banking scene, using various sources of data and time periods. Overall, these studies conclude that between 1974 and now, the banking sector has experienced substantial consolidation: the number of organizations declined significantly, while total assets in the industry increased. But asset growth was not evenly distributed among firms, rather the largest banks became significantly more concentrated.[1] The way of doing business changed, with a rise in off-balance-sheet activities, a decrease in business lending, increase in residential and business mortgage lending, and a higher price for funds, due to the loss of monopsony power over depositors. Evidence is presented to document the origins of the "too-big-to-fail" doctrine: since the 1980s, the source of the problems in the banking sector was the increased risk taking by large banks that were relatively unrestricted by the existing interstate restrictions.

The unit of observation in this paper is a BHC, a company that owns or controls one or more banks. The Gramm-Leach-Bliley Act of 1999 eliminated barriers which forced separation between commercial banks and insurance companies; it allowed BHCs to engage in financial activities, including securities underwriting and dealing, insurance agency and underwriting activities, and merchant banking activities. The Act allows existing bank holding companies to acquire full-service securities firms and insurance companies. In addition, the holding-company structure offers other attractive features: BHCs can assume shareholders' debt on a tax free basis, borrow money, acquire other banks and non-bank entities more easily, and issue stock with greater regulatory ease. They have to, however, respond to additional regulatory authorities. Klein and Saidenberg (1999) document that, on average, BHCs are better diversified, do significantly more lending, and hold significantly less capital, than their counterpart banks not part of a holding company.

This paper examines changes in the performance of BHCs over time, between 1988 and 2010. I employ the nonparametric, unconditional, hyperbolic α-quantile estimator developed by Wheelock and Wilson (2008) to estimate efficiency, that I then use to construct a hyperbolic version of the

[1] Jones and Critchfield (2005) document the differences in consolidation: the asset share of organizations with more than \$10 billion in assets increased from 42 percent in 1984 to 73 percent in 2003. The same concentration pattern exists in terms of deposits: in 2003 three organizations were holding 25 percent of the total deposits, while in 1984 there were 42 organizations holding a quarter of the nation's deposits.

Malmquist index. I decompose the index in efficiency change and technology change to determine the sources of changes at the industry level. The estimator employed presents several advantages over the traditional parametric estimators: i) it measures efficiency along a hyperbolic path, so one is not confronted with the puzzle of differences in estimates depending on the direction chosen; ii) it is a partial estimator, thus only performance of firms in the neighborhood of the firm of interest is relevant; iii) its convergence rate is comparable to that of parametric estimators ($n^{1/2}$); and iv) it is robust to outliers.

Different benchmarks are used to estimate efficiency; the three efficiency estimated are: technical, cost and revenue efficiency. Technical efficiency refers to the ability to produce the most output with a given amount of inputs. A firm is technically efficient if it is not possible to increase output without increasing input usage. Cost efficiency takes into account input prices; it shows the extent to which a firm minimizes its costs of producing some output, given the input prices that it faces. Revenue efficiency refers to the maximum feasible revenue that can be generated from some inputs, held constant, given market output prices.

Data presented in Table 3.1 suggest strong expansion has taken place: the number of BHCs has been growing every year since 1986, with the exception of 1994, when the number of BHC organizations declined by 17 percent. There is evidence of consolidation too. The size of the BHC sector, as measured by total assets, has increased consistently, in each year since 1986 to 2005; in 1994, there was an increase in the total assets held by all operating BHCs of 9 percent, despite the decline in the number of organizations, and much higher than the prior years' average growth of about 2 to 3 percent.[2] Total assets held by BHCs were \$3,192.74 billion in 1986, and \$16,413.15 billion in 2005 (values in 2005 constant dollars), an increase higher than five-fold. Other changes in the BHCs sector appear to be similar to the developments in the commercial banking sector: i) the share of total loans in total assets has increased, while the share of securities has decreased during the 1992-1995 period, ii) real estate lending has increased over time, while commercial and industrial loans have decreased, iii) the share of deposits in total liabilities declines in favor of federal funds, and iv) equity-to-assets ratio increased in 1994 (due to regulatory changes), but it continued to decline (exception is the 1999-2000 dotcom bubble period) to historically low levels until 2005.

Changes in performance between 1988 and 1998, between 1988 and 2005, and 1998 and

[2]The explanation for the dramatic consolidation in 1994 is the passing of the Riegle-Neal Act, that allowed interstate banking.

2005 are examined. Also, changes relative to 2010 are examined, but the findings are interpreted with a grain of salt, given the uncertainty about the true market value of many of assets held by BHCs. These specific years were chosen to be analyzed for several reasons. Data availability ranges from 1986 to present. The year 1988 was chosen since most banking institutions decided to write-off most of their nonperforming international loans in 1987. However, the savings and loan (S&L) crisis that affected the U.S. banking sector between 1980 and 1991, peaked between 1986 and 1990. More than 1,300 banks failed during the S&L crisis, thus 1988 was a year when the banking sector faced struggles to survive and, thus, (we assume) attempted to reorganize its production and change its business model.[3]

The year 1998 was chosen because it was the last year before the Gramm-Leach-Bliley Act was passed. This piece of legislation was an incentive for banking institutions that desired to act in the insurance market to acquire the BHC status. Finally, 2005 was chosen for two reasons. First, the financial crisis had not yet started, so one does not have to face the difficulty of pricing many of the assets on the balance sheets after 2007. While accounting documents reflected the book value at which they were acquired, markets were not able to provide information about their true market values. Second, starting June 30th 2006, only BHCs with assets higher than $500 million have to submit their quarterly reports, while the ones with lower assets had only to submit reports twice a year, and not that detailed.

The rest of the paper is organized in the following way: a short literature review is presented in Section 3.2, the estimation procedure is detailed in Section 3.3, data and empirical results are presented in Section 3.4, and Section 3.5 concludes.

3.2 Literature Review

Berger at al. (1995) analyze in great detail trends for all federally insured banking institutions between 1974 and 1994. They document the expansion of the industry, as measured by the size of total assets from $3.26 trillion in 1974 to $4.02 trillion (1994 dollars), despite the reduction in the number of banks from 12,463 to 7,926 (almost the entire reduction comes from the reduction in the number of small banks). Regulatory changes triggered a transformation in the way the banking

[3]Boyd and Gertler (1993) point to the fact that this number is misleading: many failed banks were large banks. They accounted for 4% of the failed banks, but for 60% of the total assets of failed banks.

of business was done. Banks faced an increase in the price of funds, due to loss of monopsony over depositors and current competition with money market mutual funds. Other trends were the increase in off-balance sheet activities and the reduction in commercial and industrial lending (a market niche that was covered by foreign banks). Most of the largest banks increased their equity to capital ratio. Technological and financial innovations led to improvements in credit analysis, electronic funds transfers, and the development of the derivatives markets.

Boyd and Gertler (1995) show that between 1950 and 1994, at US commercial banks there is an increase in loans, and a decline in cash and securities, due to the development of the money markets. It appears that the share of each category of loans in banks' portfolios has been unchanged between 1952 and1973. On the liabilities side, they find that the importance of checkable deposits diminished, in favor of interest-bearing liabilities: in the 1960's, the share of checkable deposits of total liabilities was 60%, and in the beginning of the 1990's, it decreased to less than 20%. They also document trends of declining equity capital-to-assets ratio until the 1980's when the regulations changed.

Stiroh (2000) examined the improved performance of BHCs between 1991 and 1997 and found that the gains were mainly due to productivity growth and changes in scale economies. Estimated cost functions showed modest economies of scale, with th largest BHCs exhibiting the strongest economies of scale. Berger et al. (1999) review the research on cost efficiency and found, based on data from the 1980s and early 1990s, that there was little evidence of efficiency improvement from mergers and acquisitions. Cost efficiency might depend on the type of merger and how the merger is implemented. Demsetz and Strahan (1997) find that large BHCs had better diversification across loan portfolios; it allowed them to operate with greater leverage and engage in more risky, and potentially, more profitable, lending without increasing firm-specific risk. Berger et al. (1996) examined a sample of banks over the period 1978 and 1990, and found evidence of small cost economies and no evidence of statistically significant revenue economies, regardless of the bank class size.

3.3 Estimation Approach

Productivity is typically defined as the amount of output obtained per unit of input used. But when the production process involves multiple inputs and outputs, productivity can not be measured reliably by simple output-input ratios. Malmquist indices are usually used in the literature to measure changes in productivity over time. A Malmquist index is defined in terms of ratios of individual firm efficiencies estimated at different points in time and against different benchmarks. Thus, in order to formally describe the Malmquist index, I have to define the efficiency concept first.

3.3.1 Efficiency Change

The economic theory defines a technology as the process through which production factors are transformed into output. With a parametric set-up, many researchers assume a functional form for the production function, usually a Cobb-Douglas function is assumed; but these functions do not have microeconomics foundations, and exhibit properties unlikely to conform to reality. The nonparametric approaches to analyzing efficiency rely on the microeconomics theory of the firm: it is assumed a production set exists and it is closed by a frontier (referred to as the production frontier or technology) that "envelops all the data observed," hence the general name of data envelopment techniques. Since these methods do not require specification of a functional relationship between variables, few other assumptions are necessary. In particular, no assumptions about the frontier shape or the distribution of inputs and outputs on the production set are necessary.

Technical efficiency refers to the ability to produce the most output with a given amount of inputs. A firm is technically efficient if it is not possible to increase output without increasing input usage, so estimating efficiency in a nonparametric framework involves comparing the individual performance of one firm with a benchmark defined by (all or some) firms in the sample. Specifically, efficiency is measured as the distance from where the firm lies in the input–output space to the production frontier (technology) that envelops the data.

I use the unconditional hyperbolic α-quantile estimator developed by Daouia and Simar (2007) Wheelock and Wilson (2008) to estimate efficiency. This estimator differs from "traditional" estimators in several respects. First, the unconditional hyperbolic α-quantile estimator is a "partial" frontier estimator, in the sense that the benchmark comparison consists of firms in the neighborhood of the analyzed firm, and the desired size of the neighborhood can be chosen by the researcher. Second, traditional estimators estimate the distance from a fixed data point to the full frontier in a direction orthogonal to the output axis (in the input orientation case), or to the input axis (in the output orientation case). If the sector analyzed operates under variable returns to scale, then the choice of input or output orientation has a big impact on the measured efficiency: a firm could lie close to the frontier in the output direction, but far from it in the input direction, while another firm could be close to the frontier in the input direction, but far from it in the output direction. The α-quantile estimator allows estimation of the efficiency along a hyperbolic path, such that inputs and outputs are adjusted simultaneously, rather than just in the input or output direction, and, thus, overcoming the issue of direction choice. Finally, this estimator exhibits, unlike traditional nonparametric estimators, desirable statistical properties: it is robust to outliers in the data and has a high convergence rate, comparable to the parametric estimators' convergence rate.

The formal discussion of the economic model and econometric estimation is based on Wheelock and Wilson (2008, 2010). Consider the following production possibilities set:

$$P \equiv \{(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x} \text{ can produce } \boldsymbol{y}\} \subset \mathbb{R}_{+}^{p+q}, \tag{3.1}$$

where $\boldsymbol{x} \in \mathbb{R}_{+}^{p}$ and $\boldsymbol{y} \in \mathbb{R}_{+}^{q}$ denote vectors of inputs and outputs, respectively, and P^{δ} denotes the upper boundary of the production set P, the technology frontier.

The goal is to estimate the distance from an observation, a point $(\boldsymbol{x}, \boldsymbol{y})$, to the frontier P^{δ}. The hyperbolic-graph distance function measures the distance from a fixed point $(\boldsymbol{x}, \boldsymbol{y})$ to P^{δ} along the hyperbolic path $(\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y})$, where $\gamma \in (0, 1)$. The hyperbolic distance function is given by

$$\gamma(\boldsymbol{x}, \boldsymbol{y} \mid P) \equiv \sup\left\{\gamma > 0 \mid (\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y}) \in P\right\}. \tag{3.2}$$

Note that P represents the true production set. We only observe a sample of iid random variables with probability density function $f(\boldsymbol{x}, \boldsymbol{y})$ with support over P. The density $f(\boldsymbol{x}, \boldsymbol{y})$ implies the following probability function

$$H(\boldsymbol{x}_0, \boldsymbol{y}_0) = Pr(\boldsymbol{x} \leq \boldsymbol{x}_0, \boldsymbol{y} \geq \boldsymbol{y}_0), \tag{3.3}$$

which gives the likelihood of drawing an observation from the $f(\boldsymbol{x}, \boldsymbol{y})$ that weakly dominates the agent operating at $(\boldsymbol{x}_0^{\delta}, \boldsymbol{y}_0^{\delta}) \in P^{\delta}$.[4] The hyperbolic α-quantile distance function can be expressed as

$$\gamma_{\alpha}(\boldsymbol{x}, \boldsymbol{y}) = \sup\left\{\gamma > 0 \mid H(\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y}) > (1-\alpha)\right\}. \tag{3.4}$$

[4] An observation $(\widetilde{x}, \widetilde{y})$ weakly dominates (x, y) if $\widetilde{x} \leq x$ and $\widetilde{y} \geq y$.

For $0 < \alpha < 1$ and a fixed point $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{p+q}$, $\gamma_\alpha(\boldsymbol{x}, \boldsymbol{y}) > 1$ gives the proportionate, simultaneous decrease in inputs and increase in outputs required to move from $(\boldsymbol{x}, \boldsymbol{y})$ along a path $(\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y})$, $\gamma > 0$, to a point with $(1-\alpha)$ probability of being weakly dominated.

Note that the above estimation does not take into account the role of prices; in this sense, technical efficiency is the most restrictive type of efficiency. It is possible for a firm to be technical efficient, but to use a mix of inputs or outputs that are not optimal, i.e., a mix that does not minimize cost, or maximize revenue, respectively. Cost and revenue efficiency are more inclusive efficiency measures. Cost efficiency reflects the extent to which a firm minimizes its costs of producing some output, given input prices faced and holding output constant. Formally, production cost is given by $\mathbf{w'}_x\mathbf{x}$, where $\mathbf{w}_x$ denotes a w vector of input prices, and the set of feasible combinations of cost and outputs is given by $C(\mathbf{w} \mid P)$. Then, cost efficiency measured in the hyperbolic direction can be estimated analogously to the technical efficiency estimation:

$$\gamma_\alpha(\mathbf{w}'\boldsymbol{x}, \boldsymbol{y} \mid C(\mathbf{w} \mid P)) \equiv \sup\left\{\gamma > 0 \mid (\gamma^{-1}\mathbf{w}'\boldsymbol{x}, \gamma\boldsymbol{y}) \in C(\mathbf{w} \mid P)\right\}. \tag{3.5}$$

For a firm facing input prices $\mathbf{w}$, $\gamma_\alpha(\mathbf{w}'\boldsymbol{x}, \boldsymbol{y} \mid C(\mathbf{w} \mid P)) > 1$ gives the proportionate, simultaneous decrease in costs and increase in outputs, required to move for a point $(\mathbf{w}'\boldsymbol{x}, \boldsymbol{y})$ to a point with $(1-\alpha)$ probability of being weakly dominated.

Revenue efficiency refers to the maximum feasible revenue that can be generated from some inputs, held constant, given market output prices. Similarly to cost efficiency, technical efficiency does not imply revenue efficiency, but revenue inefficiency does imply technical inefficiency. Revenue is given by $\mathbf{z'}_y\mathbf{y}$, where $\mathbf{z}_y$ denotes a q vector of output prices, and the set of feasible combinations of inputs and revenues is given by $R(\mathbf{z} \mid P)$. The estimation of the revenue efficiency in the hyperbolic direction is given by:

$$\gamma_\alpha(\boldsymbol{x}, \mathbf{z}'\boldsymbol{y} \mid R(\mathbf{z} \mid P)) \equiv \sup\left\{\gamma > 0 \mid (\gamma^{-1}\boldsymbol{x}, \gamma\mathbf{z}'\boldsymbol{y}) \in R(\mathbf{z} \mid P)\right\}. \tag{3.6}$$

For a firm facing output prices $\mathbf{z}$, $\gamma_\alpha(\boldsymbol{x}, \mathbf{z}'\boldsymbol{y} \mid R(\mathbf{z} \mid P)) > 1$ gives the proportionate, simultaneous decrease in inputs and increase in revenue, required to move from a point $(\boldsymbol{x}, \mathbf{z}'\boldsymbol{y})$ to a point with $(1-\alpha)$ probability of being weakly dominated.

3.3.2 Productivity Change

The Malmquist index measures total factor productivity change over time. The concept is based on Malmquist (1953)'s idea of using distance functions to develop an index for productivity change. I define the Malmquist index in terms of the hyperbolic α-quantile measure of efficiency. Let P_α^t denote the production set at time t for the firms comprised in the set defined by the specified α (production set defined analogously to the production set defined in (3.1)), and $P_\alpha^{\delta t}$ denote the upper boundary of the production set P_α^t. The index is then given by:

$$\mathcal{M}_{\alpha,i}(t_1,t_2) \equiv \left[\frac{\gamma(\mathbf{x}_{it_2},\mathbf{y}_{it_2}) \mid \mathcal{V}(P_\alpha^{t_1}))}{\gamma(\mathbf{x}_{it_1},\mathbf{y}_{it_1}) \mid \mathcal{V}(P_\alpha^{t_1}))} \times \frac{\gamma(\mathbf{x}_{it_2},\mathbf{y}_{it_2}) \mid \mathcal{V}(P_\alpha^{t_2}))}{\gamma(\mathbf{x}_{it_1},\mathbf{y}_{it_1}) \mid \mathcal{V}(P_\alpha^{t_2}))}\right]^{\frac{1}{2}}. \tag{3.7}$$

The term $\gamma_\alpha(\mathbf{x}_{it_k},\mathbf{y}_{it_k}) \mid \mathcal{V}(P^{t_j}))$ is an estimate of the distance from firm i's location at time t_k to the boundary of $\mathcal{V}(P^{t_j})$, where $\mathcal{V}(P^{t_j})$ represents the convex cone of the hyperbolic α production set $(P_\alpha^{t_j})$. Fare and Grosskopf (1996) and Ball et al. (2005) showed that if the technology allows for variable returns to scale, the index ignores the contribution of scale economies to productivity growth. In order for the Malmquist index to indicate the true total factor productivity changes, the index must be defined in terms of constant returns to scale, thus the convex cone is used rather than the convex hull. With variable returns to scale, technically efficient firms operating along the increasing or decreasing returns regions of the technology are less productive than the technically efficient firms operating along the constant returns region of the frontier. The Malmquist index is the geometric mean of two ratios that measure the change in productivity using as a benchmark the convex cone of the set bounded by the technology prevailing at time t_1 and t_2, respectively.

We can decompose the hyperbolic-quantile-based Malmquist index in a measure of α-quantile-based efficiency change:

$$\mathcal{E}_{\alpha,i}(t_1,t_2) = \frac{\gamma_\alpha^{t_2}(\mathbf{x}_{it_2},\mathbf{y}_{it_2})}{\gamma_\alpha^{t_1}(\mathbf{x}_{it_1},\mathbf{y}_{it_1})}, \tag{3.8}$$

and a measure of the industry-wide technology change:

$$\mathcal{T}_{\alpha,i}(t_1,t_2) \equiv \left[\frac{\gamma_\alpha^{t_1}(\mathbf{x}_{it_1},\mathbf{y}_{it_1})}{\gamma_\alpha^{t_2}(\mathbf{x}_{it_1},\mathbf{y}_{it_1})} \times \frac{\gamma_\alpha^{t_1}(\mathbf{x}_{it_2},\mathbf{y}_{it_2})}{\gamma_\alpha^{t_2}(\mathbf{x}_{it_2},\mathbf{y}_{it_2})}\right]^{\frac{1}{2}}. \tag{3.9}$$

$\mathcal{E}_{\alpha,i}(t_1,t_2)$ measures the change in efficiency between times t_1 and t_2, relative to the hyperbolic

α-quantile frontiers at times t_1 and t_2. An estimate lower than 1 shows an increase in technical efficiency measured relative to the α-quantiles at times t_1 and t_2, respectively; a value higher than 1 shows a decrease in technical efficiency. $\mathcal{T}_{\alpha,i}(t_1, t_2)$ is the geometric mean of two ratios that measure the shift in the α-quantile frontier relative to a firm's position at times t_1 and t_2. $\mathcal{T}_{\alpha,i}(t_1, t_2) < 1$ indicates that the α-quantile has shifted upwards, while a figure higher than 1 is indicative of the downward shift of the frontier.

As one can define different efficiency measures, depending of the benchmark used. Ball et al. (2005) and Wheelock and Wilson (2010) defined, similarly, different Malmquit indices and their decompositions using various benchmarks. If a cost frontier is used, then the Malmquist cost productivity can be defined as:

$$\mathcal{MC}_{\alpha,i}(t_1, t_2) \equiv \left[\frac{\gamma(\mathbf{w}'_{it_2}\mathbf{x}_{it_2}, \mathbf{y}_{it_2}) \mid \mathcal{V}(C_\alpha(\mathbf{w}_{it_1}, P^{t_1})))}{\gamma(\mathbf{w}'_{it_1}\mathbf{x}_{it_1}, \mathbf{y}_{it_1}) \mid \mathcal{V}(C_\alpha(\mathbf{w}_{it_1}, P^{t_1})))} \times \frac{\gamma(\mathbf{w}'_{it_2}\mathbf{x}_{it_2}, \mathbf{y}_{it_2}) \mid \mathcal{V}(C_\alpha(\mathbf{w}_{it_2}, P^{t_2})))}{\gamma(\mathbf{w}'_{it_1}\mathbf{x}_{it_1}, \mathbf{y}_{it_1}) \mid \mathcal{V}(C_\alpha(\mathbf{w}_{it_2}, P^{t_2})))}\right]^{\frac{1}{2}}, \tag{3.10}$$

which can be decomposed into a measure of cost-efficiency change:

$$\mathcal{EC}_{\alpha,i}(t_1, t_2) = \frac{\gamma_\alpha^{t_2}(\mathbf{w}'_{it_2}\mathbf{x}_{it_2}, \mathbf{y}_{it_2})}{\gamma_\alpha^{t_1}(\mathbf{w}'_{it_1}\mathbf{x}_{it_1}, \mathbf{y}_{it_1})}, \tag{3.11}$$

and a measure of industry-wide cost technology change:

$$\mathcal{TC}_{\alpha,i}(t_1, t_2) \equiv \left[\frac{\gamma_\alpha^{t_1}(\mathbf{w}'_{it_1}\mathbf{x}_{it_1}, \mathbf{y}_{it_1})}{\gamma_\alpha^{t_2}(\mathbf{w}'_{it_1}\mathbf{x}_{it_1}, \mathbf{y}_{it_1})} \times \frac{\gamma_\alpha^{t_2}(\mathbf{w}'_{it_2}\mathbf{x}_{it_2}, \mathbf{y}_{it_2})}{\gamma_\alpha^{t_2}(\mathbf{w}'_{it_2}\mathbf{x}_{it_2}, \mathbf{y}_{it_2})}\right]^{\frac{1}{2}}. \tag{3.12}$$

When a revenue frontier is employed, the Malmquist revenue productivity is defined as:

$$\mathcal{MR}_{\alpha,i}(t_1, t_2) \equiv \left[\frac{\gamma(\mathbf{x}_{it_2}, \mathbf{z}'_{it_2}\mathbf{y}_{it_2}) \mid \mathcal{V}(R_\alpha(\mathbf{z}_{it_1}, P^{t_1})))}{\gamma(\mathbf{x}_{it_1}, \mathbf{z}'_{it_1}\mathbf{y}_{it_1}) \mid \mathcal{V}(R_\alpha(\mathbf{z}_{it_1}, P^{t_1})))} \times \frac{\gamma(\mathbf{x}_{it_2}, \mathbf{z}'_{it_2}\mathbf{y}_{it_2}) \mid \mathcal{V}(R_\alpha(\mathbf{z}_{it_2}, P^{t_2})))}{\gamma(\mathbf{x}_{it_1}, \mathbf{z}'_{it_1}\mathbf{y}_{it_1}) \mid \mathcal{V}(R_\alpha(\mathbf{z}_{it_2}, P^{t_2})))}\right]^{\frac{1}{2}}, \tag{3.13}$$

the revenue-efficiency change measure as:

$$\mathcal{ER}_{\alpha,i}(t_1, t_2) = \frac{\gamma_\alpha^{t_2}(\mathbf{x}_{it_2}, \mathbf{z}'_{it_2}\mathbf{y}_{it_2})}{\gamma_\alpha^{t_1}(\mathbf{x}_{it_1}, \mathbf{z}'_{it_1}\mathbf{y}_{it_1})}, \tag{3.14}$$

and the industry-wide revenue technology change as:

$$\mathcal{TR}_{\alpha,i}(t_1,t_2) \equiv \left[\frac{\gamma_\alpha^{t_1}(\mathbf{x}_{it_1},\mathbf{z}'_{it_1}\mathbf{y}_{it_1})}{\gamma_\alpha^{t_2}(\mathbf{x}_{it_1},\mathbf{z}'_{it_1}\mathbf{y}_{it_1})} \times \frac{\gamma_\alpha^{t_2}(\mathbf{x}_{it_2},\mathbf{z}'_{it_2}\mathbf{y}_{it_2})}{\gamma_\alpha^{t_2}(\mathbf{x}_{it_2},\mathbf{z}'_{it_2}\mathbf{y}_{it_2})}\right]^{\frac{1}{2}}. \tag{3.15}$$

The measures given by (3.10) and (3.13) have the same qualitative interpretation as the measure in (3.7), with the only difference that the benchmark is now given by $\mathcal{V}(C_\alpha(\mathbf{w}^{\mathbf{t_j}}) \mid P^{t_j})$ and $\mathcal{V}(R_\alpha(\mathbf{z}^{\mathbf{t_j}}) \mid P^{t_j})$, respectively, and not by $\mathcal{V}(P_\alpha^{t_j})$. Similarly, the measures in (3.11) and (3.14) are to be interpreted in a similar manner to the measure in (3.8), and the measures in (3.12) and (3.15) similarly to the measure in (3.9).

3.4 Data and Results

3.4.1 Data

Data used in this paper are first quarter data from the FR 9Y-C forms that BHCs have to submit quarterly to the Federal Reserve. Table 3.2 gives the definitions for the variables used as inputs and outputs. For the technical efficiency and technology change, the following inputs were used: labor, purchased funds, core deposits, physical capital, and equity capital. The considered outputs were: real estate loans, business loans, consumer loans, other loans, securities, and off-balance items. The choice of inputs/ outputs is in line with the current literature and is consistent with the intermediation approach. Except labor, which is reported as number of full-time equivalent employees, all inputs and outputs are in dollar figures. For the cost efficiency and cost technology change, the cost variable was computed as the sum of inputs multiplied by their respective prices. Note that no price is available for equity capital, thus equity was not considered in this set of estimations. For the estimations that used as benchmark the revenue frontier, revenue is calculated as the sum of securities and loans multiplied by their respective average prices. The off-balance items output was not considered, for the same reason of not being able to identify a price for it.

Table 3.3 gives the summary statistics for the inputs and outputs used, for each of the quarters examined. Figures are in thousands of 2005 constant dollars. Observations that had missing or negative values for some inputs or outputs were dropped. The sample contains BHCs that are very different, both in terms of size and activity. The median value is much lower than the mean value for each variable, in each quarter. Also, the minimum and maximum values reflect great size differences. In terms of assets and liabilities, there are BHCs that do not lend in all markets, and there are BHCs that do not attract funds from all potential savers.

Table 3.4 reports the number of BHCs in the sample, by year and by period. The sample is not balanced, and less than 20% of the BHCs operating in 1988 were still operating in 2010: out of the 1,323 BHCs operating in 1988, 575 were still in business in 1998, 432 in 2005, and only 262 in 2010. About 38% out of the BHCs operating in 1998 are present in our 2010 sample, and 37% out of the BHCs operating in 2005. These figures are indicative of the consolidation trend in the industry (though the decline in numbers for 2010 are an artifact of the changes in the reporting regulations).

3.4.2 Results

Tables 3.5–3.10 report estimates of the geometric mean changes of efficiencies and technologies by quartiles of total assets, where Q1 is the first quartile, and comprises the smallest 25% of the total sample. The geometric, rather than the arithmetic mean is more appropriate, given the multiplicative nature of the estimates. Each cell reports two numbers: the top one is the (geometric) mean of the estimates for the observations that belonged to Qi in t_1 and Qj in t_2. The reported results are obtained for a value of $\alpha = 0.99$ The bottom figure gives the number of BHCs in those respective quartiles in the examined time periods. For instance, there were 48 BHCs that were among the smallest 25% (Q1) in both 1988 and 1998, 29 BHCs that were among the smallest 25% in 1988, but among the second-smallest (Q2) in 1998, 17 BHCs that were among the smallest 25% in 1988, but among the third-smallest (Q3) in 1998, and only one that was among the smallest in 1988, and among the largest (Q4) in 1998.[5]

Table 3.5 reports the point estimates for the mean changes in efficiency for all the time periods examined. I reiterate that the estimates pertaining to 2010 have to be interpreted with caution, given the uncertainty about the true market value of some assets on the BHCs' balance sheets. The results suggest that most BHCs experienced a decline in mean efficiency between 1988 and 1998. BHCs that were in the first quartile in both 1988 and 1998, experienced a mean average

[5]The ranges of quartiles for each quarter are as following (in thousands of constant 2005 $): for 1988, (0–174,916), (175,111-322,717), (323,387–799,782), and (800,997–3.09e+08); for 1998, (0–251,421), (251,605–387,041), (387,291–900,654), and (905,441–4.32e+08), for 2005, (0–224,454), (224,456–350,069), (350,190–692,398), and (693,495-1.52e+09); and for 2010, (0–569,968), (572,854–848,887), (850,180–1,722,122) and (1,726,862–2.14e+09).

decrease of about 3.1%, while those that were in the second largest quartile (Q3) in 1988 and moved to the largest quartile (Q4) in 1998 experienced a mean decrease in efficiency of about 11.5%.

The BHCs that were among the largest in each of the two years experienced a mean increase in efficiency of about 15.2%. The trend for the largest 1988 BHCs that were among the largest ones in 2005 and 2010 was the same: an improvement in mean efficiency of 45% and 13% respectively. An interesting pattern exists for the BHCs that belonged to the smallest quartile in 1988, but then grew over time. With one exception, all their mean efficiencies appear to have worsened, with the largest decline for banks that grew extensively, belonging to the largest class size in 2005 or 2010.

Estimates of the mean efficiency change between 1998 and 2005 presented in Table 3.5 suggest that, on average, all BHCs, in all class sizes, became more efficient, with the largest gains for the firms that ended in the largest quartile in 2005. For the 1998–2010 time period, besides the increase in average efficiency for the BHCs that ended up in the largest quartile, also BHCs that ended up in the smallest quartile in 2010 seem to have experienced an increase in mean efficiency. The results for the mean average changes between 2005 and 2010 show that BHCs that belonged to the smallest or second smallest quartile in 2010 had, on average, improved their technical efficiency. These results suggest that until 2005 (so before the onset of the current financial crisis), the largest BHCs, or those that became larger over time, experienced, on average, improvements in technical efficiencies. Also, it appears that there is no evidence of a "catching-up" effect, in the sense that there are gains in performance due to large size: BHCs that were the largest in the beginning of one period, and still belonged to the largest class size by the end of the period, consistently improved their performance, while small BHCs that increased their size experienced worsened performance.

Table 3.6 reports estimates of the mean technology change. Estimates higher than 1 indicate an downward shift of the technology (so a reduced production possibilities frontier), while values lower than 1 reflect an upward shift of the frontier. Interestingly, for all time period examined, except the period of 2005-2010, all BHCs, regardless of class size, experienced an downward shift of the technology. For the 2005–2010 time period, BHCs that belonged to the two largest quartiles in 2005 experienced an upward shift of the frontier.

Cost efficiency changes estimates are presented in Table 3.7. These estimates are qualitatively similar to the technical efficiency ones. BHCs that were the largest in 1988 and continued growing, such that they were part of the fourth quartile subsequently, experienced a mean improvement in cost efficiency over time: they were 13.8% more cost efficient in 1998, 43.5% more cost efficient in 2005, and 29.9% in 2010. BHCs that were among the smallest in 1988, but increased their size over time, became, on average, less cost efficient. Estimates for the 1998–2005 period indicate that BHCs in all class sizes improved their cost efficiency, while between 1998 and 2010 all of them, except the ones in the largest two quartiles, experienced declines in cost efficiencies. Cost efficiency declined on average between 2005 and 2010 for all BHCs, except for the ones in the smallest class size. BHCs belonging to the fourth quartile in each year improved their cost efficiency in all time intervals examined, except for 2005-2010 period, when they experienced a decrease in mean cost efficiency of 18.2%.

Table 3.8 presents the estimates for the cost technology changes. There is evidence of an upward shift in the cost frontier for almost all class sizes, for each time period, indicating an increase in the cost of producing given amounts of output. Exceptions are the largest BHCs in 1988 − 2005 and 1998–2005 period, and the smallest BHCs during 2005-2010, that experienced an downward shift of the α-cost frontier. Put together, the information in Tables 6 and 7 reveals several things. First, for the time intervals 1988–1998, 1988–2005, 1988–2010, and 1998–2005, I find: i) most BHCs became more cost-efficient and experienced an upward shift of the cost α-quantile, suggesting that while most of them faced higher production costs, shifts in technology reduced their distance from the frontier, thus the explanation for the increase in cost-efficiency; ii) the smallest BHCs, that belonged to the first quartile in all years became more cost-inefficient, and experienced an upward shift of the cost frontier, an indication of worsened productivity for these BHCs over the time periods mentioned; and iii) the group of the largest BHCs became more cost-efficient and experienced a downward shift in the partial cost frontier from 1988 to 2005 and 1998 to 2005, a sign of improved cost–productivity over the time frames mentioned. Second, over the period 1998 and 2010, almost all BHCs experienced a decline in productivity, triggered by a simultaneous decrease in mean efficiency and an upward shift of the cost frontier, while the large BHCs experienced higher production costs as well, but they became, on average, more cost-efficient. Ultimately, between 2005 and 2010, all BHCs experienced a decline in cost– productivity, indicated by both decrease in efficiency and increase in minimum production costs.

Table 3.9 reports the mean revenue efficiency changes. BHCs belonging to the fourth quartile in each year, experienced positive changes in mean revenue efficiency, except for the time period of 2005 to 2010, when their revenue efficiency declined by 8.7%. BHCs of medium size (so those that belonged to either second or third quartile) show a worsening, on average, of their revenue efficiency. The findings for the mean changes in revenue technology, reported in Table 10, show that the α-quantile revenue frontier has shifted downward for all time periods, all class sizes, with two exceptions: the BHCs that belonged to the largest quartile in both 2005 and 2010, and the ones that belonged to the first quartile in 1998 and 2005. A downward shift of the revenue α-quantile indicates that less revenue is generated from the output obtained with some given amount of input. All six periods examined exhibit a similar pattern: the smaller the class size BHCs belong to, the more dramatic the shift in the revenue frontier.

Combining the information from Tables3. 9 and 3.10, the findings suggest that medium-size BHCs worsened their revenue-productivity over time, since they became less revenue-efficient and faced a downward shift of the revenue frontier, for all time periods examined. Almost all smallest and largest BHCs became more revenue-efficient, on average, but gained lower revenues (as evidenced by the downward shift of the revenue frontier). Hence, the improvement in revenue-efficiency could be attributed to the shift in the technology, that reduced their distance from the frontier.

3.5 Conclusion

The banking business environment has changed over the last decades significantly, due to changes triggered by new legislation, or advancements in technology and finance. This paper examines the performance of BHCs between 1988 and 2010 period, by analyzing mean changes in technical, cost, and revenue efficiency, and mean changes in technology, cost technology and revenue technology. I find that, over the years, the largest banks experienced the largest gains, in technical, cost and revenue efficiency, with the exception of 2005 –2010 period, when the smallest BHCs seem to have experienced gains in all efficiencies. Estimates of the technology change show an downward shift of the α-quantile (i.e., a decrease in the output produced for some given input used), an upward shift of the cost α-quantile (i.e., an increase in the minimum cost of producing some given output), and downward shift of the revenue α-quantile (i.e., a decrease in the amount of revenue generated from the output produced with some given amount of input), for most periods and class size, except the large BHCs. These results indicate that the largest BHCs have improved their performance over 1988–2005 period, while the other BHCs experienced a worsening of their performance over time.

It is interesting to note that only the BHCs that belonged to the largest quartile appear to have the highest gains. Over the time periods examined, the size of the BHCs had increased consistently. These findings are in line with the strand of literature that suggests that the current legislation is conducive to consolidation, and to the strand of research documenting that large banking institutions benefited more from the technological improvement and financial market globalization. The findings are mixed for the small BHCs, and the medium-size BHCs appear to have had negative changes in their efficiencies and technologies. These results support the literature that finds that consolidation leads to benefits from greater diversification that are, in turn, offset by the costs of increased risk-taking.

Table 3.1: Number of BHCs over Time

Year	Total number of BHCs	BHCs with $500 mill. or more in assets	Total assets (values in $ bill.)
1986	1,161	364	3,912.74
1987	1,345	352	3,933.03
1988	1,385	361	4,030.52
1989	1,423	372	4,036.71
1990	1,565	493	5,097.86
1991	1,589	506	4,923.58
1992	1,624	511	5,022.93
1993	1,616	476	5,148.76
1994	1,338	479	5,631.78
1995	1,342	499	5,940.39
1996	1,397	513	6,105.33
1997	1,456	537	6,435.55
1998	1,512	526	6,998.07
1999	1,643	546	8,257.72
2000	1,718	594	9,750.58
2001	1,800	628	10,814.77
2002	1,915	662	11,330.89
2003	2,059	718	12,210.92
2004	2,212	756	14,207.65
2005	2,294	792	16,413.15
2006	987	829	11,166.38
2007	973	862	11,672.89
2008	959	881	10,959.74
2009	999	946	13,996.48

NOTES:
1. Data for the first quarter in each respective year, except 1986 for which the information pertains to the third quarter.
2. Data made publicly available by the Federal Reserve are based on the reports BHCs have to submit quarterly. Since not all the BHCs have to report, these figures underreport the true number of BHCs. The Report to the Congress on Financial Holding Companies under the Gramm-Leach-Bliley Act (2003) states, for instance, that the total number of operating BHCs in December 2000 was 5,072, 5,090 in December 2001, and 5,094 in December 2002. These figures are approximately three times higher than the total BHCs number from the publicly available data.
3. Data in constant 2005 dollars, deflated using the GDP Implicit Price Deflator.
4. The decline in the BHCs number reflects changes in regulation: starting 2006, the Federal Reserve changed the requirements for the reporting BHCs: from a prior threshold of $150 million in assets, starting with 2006, the threshold was increased to $500 million. Thus, though it is possible that the number of BHCs decreased slightly too.

Table 3.2: Variables Definitions

Labor – number of full-time equivalent employees
Purchased funds – time deposits over $100,000, federal funds purchased in domestic offices, securities sold under agreement to repurchase, trading liabilities, other borrowed money, subordinated notes and debentures, subordinated notes payable to unconsolidated trusts issuing trust preferred securities, and trust preferred securities issued by consolidated special purpose entities.
Core deposits – domestic transactions accounts, time deposits under $ 100,000, and savings deposits
Physical capital – Premises and fixed assets (including capitalized leases)
Equity capital – total equity capital plus noncontrolling (minority) interests in consolidated subsidiaries
Labor price – salaries and employee benefits/ labor
Purchased funds price – interest expense on time deposits of $100,000 or more, expense on federal funds purchased and securities sold under agreements to repurchase, interest expense on trading liabilities and other borrowed money, interest expense on subordinated notes and debentures and on mandatory convertible securities / stock of purchased funds
Core deposits price – total interest on time deposits less than $ 100,000, plus interest on other deposits/ stock of core deposits
Physical capital price – Expenses of premises and fixed assets, net of rental income (excluding salaries and employee benefits and mortgage interest)/ Premises and fixed assets

Real estate loans – loans secured by real estate
Consumer loans – loans to individuals for household, family, and other personal expenditures (includes purchased paper): credit cards, credit plans, other consumer loans
Business loans – commercial and industrial loans to US and non-US addressees
Other loans – total loans less real estate loans, business loans and consumer loans
Securities – held-to-maturity securities, available-for-sale securities, federal funds sold in domestic offices, securities purchased under agreements to resell
Off-balance items – noninterest income minus service charges
Loans price – interest and fee income on loans in domestic offices/ total loans
Securities price – gains (losses) from held-to-maturity securities, plus gains (losses) from available-for sale securities, plus interest income from federal funds sold and securities purchased under agreements to resell/ stock of securities

Variable cost=(labor x labor price) + (purchased funds x purchased funds price) + (core deposits x core deposits price) + (physical capital x physical capital price)
Revenue=[(real estate loans+business loans +consumer loans+other loans) x loans price] + (securities x securities price)

Table 3.3: Summary Statistics

Variable	Mean	Median	St. Dev.	Min.	Max.
1988 n=(1,323)					
cost	49,703	5,503	240,400	227	6,160,488
revenue	42,957	5,116	182,084	85	4,007,057
labor	1,052	138	4,347	7	90,000
purchased funds	744,801	44,613	3,884,490	0	88,036,536
core deposits	1,491,171	245,360	5,146,421	12,275	73,950,208
physical capital	48,311	6,121	219,100	0	5,061,705
equity capital	170,559	25,368	659,790	119	13,600,000
real estate loans	626,428	86,715	3,025,963	380	78,855,160
business loans	636,812	44,748	3,197,323	0	55,630,060
consumer loans	354,435	29,598	1,972,110	0	57,241,708
other loans	301,562	12,469	1,794,726	0	28,335,506
off-balance items	10,081	309	71,473	0	1,909,934
securities	564,284	94,297	2,109,004	5,281	37,701,704
1998 (n=1,502)					
cost	60,738	5,341	331,560	441	5,251,706
revenue	56,319	5,552	283,537	321	4,703,494
labor	1,287	146	6,528	13	106,240
purchased funds	1,411,068	58,320	10,074,227	0	214,786,432
core deposits	2,195,308	282,254	10,366,007	0	171,007,584
physical capital	65,576	7,436	354,731	0	5,363,821
equity capital	395,330	37,544	2,093,045	766	33,722,732
real estate loans	1,180,320	153,139	5,603,618	0	94,699,016
business loans	757,000	38,333	4,619,897	0	78,369,608
consumer loans	483,723	26,100	3,050,176	0	65,750,176
other loans	293,276	8,247	2,006,102	0	36,980,268
off-balance items	23,568	433	157,420	5	2,814,678
securities	1,046,803	115,173	5,924,996	5,280	111,621,088

Table 3.3 – continued

Variable	Mean	Median	St. Dev.	Min.	Max.
		2005 (n=2,283)			
cost	61,704	3,218	575,865	226	1.35e+07
revenue	54,854	3,899	468,871	59	9,283,821
labor	1,446	111	13,300	12	315,935
purchased funds	2,617,033	79,003	27,082,735	1,197	642,884,928
core deposits	2,539,293	231,704	21,730,681	6,916	520,162,240
physical capital	62,236	6,766	465,942	73	9,504,360
equity capital	647,755	30,993	6,056,327	5,311	120,564,552
real estate loans	2,010,958	177,296	15,962,526	0	311,206,912
business loans	679,075	33,712	6,187,991	0	118,629,544
consumer loans	679,542	11,022	8,134,545	0	179,799,616
other loans	324,438	5,874	3,141,593	0	70,143,480
off-balance items	42,129	375	446,441	2	11,264,050
securities	1,791,119	79,425	17,900,000	0	414,107,904
		2010 (n=958)			
cost	102,795	6,847	796,225	955	1.47e+07
revenue	98,870	8,343	808,663	709	1.42e+07
labor	2,370	228	17,856	29	289,070
purchased funds	6,365,284	222,182	58,130,366	0	980,933,568
core deposits	5,156,111	539,747	38,588,162	42,409	751,416,384
physical capital	118,694	16,482	735,249	157	13,839,608
equity capital	1,386,006	74,927	10,816,192	256	211,541,744
real estate loans	3,456,057	446,150	26,030,615	0	486,080,640
business loans	1,083,046	75,148	8,113,177	40	150,175,008
consumer loans	1,325,661	16,179	13,628,897	0	253,857,904
other loans	621,339	14,930	5,193,127	0	87,309,480
off-balance items	103,376	1,135	896,824	5	14,621,803
securities	4,304,046	170,793	36,797,850	548	636,865,472

Table 3.4: Number of BHCs in the Sample

1988	**1998**	**2005**	**2010**
1,323	1,502	2,283	958
1988–1998	**1988–2005**	**1988–2010**	
575	432	262	
1998–2005	**1998–2010**		
1,013	584		
2005–2010			
850			

Table 3.5: Mean Estimates of Efficiency Change by Quartile

1988-1998

1988	1998 Q1	Q2	Q3	Q4
Q1	1.031	1.062	1.074	-
	48	29	17	1
Q2	1.001	0.988	1.005	0.928
	21	41	93	18
Q3	0.840	1.088	1.003	1.115
	3	7	76	68
Q4	1.803	2.588	1.577	0.848
	1	1	2	150

1988-2005

1988	2005 Q1	Q2	Q3	Q4
Q1	1.054	1.113	1.011	1.068
	23	39	34	12
Q2	0.959	0.981	0.927	0.867
	3	16	45	63
Q3	-	-	1.005	0.818
	0	0	13	90
Q4	-	-	-	0.550
	0	0	0	94

1988-2010

1988	2010 Q1	Q2	Q3	Q4
Q1	1.035	0.980	1.121	2.144
	11	11	9	2
Q2	0.883	0.965	1.059	0.888
	17	18	32	16
Q3	1.011	1.215	0.965	0.870
	3	6	36	32
Q4	-	-	1.096	0.677
	0	0	1	70

Table 3.5 – continued

1998-2005

1998	2005 Q1	Q2	Q3	Q4
Q1	0.976	0.962	0.946	0.792
	42	144	84	6
Q2	1.139	0.945	0.919	0.898
	2	53	172	41
Q3	-	1.273	0.961	0.879
	0	2	73	185
Q4	-	-	-	0.726
	0	0	0	209

1998-2010

1998	2010 Q1	Q2	Q3	Q4
Q1	0.959	1.011	0.917	0.877
	43	21	16	2
Q2	0.916	1.011	0.954	0.881
	61	61	30	7
Q3	0.976	1.025	1.028	0.892
	15	44	107	36
Q4	-	1.231	1.100	0.852
	0	1	8	132

2005-2010

2005	2010 Q1	Q2	Q3	Q4
Q1	0.825	0.812	-	-
	13	2	0	0
Q2	0.958	1.142	1.161	-
	56	14	6	0
Q3	1.015	1.092	1.059	0.996
	141	166	64	2
Q4	0.565	1.207	1.130	1.168
	2	25	149	210

Table 3.6: Mean Estimates of Technology Change by Quartile

1988-1998

1988	1998 Q1	Q2	Q3	Q4
Q1	1.233	1.141	1.096	1.230
	48	29	17	1
Q2	1.045	1.040	1.032	1.027
	21	41	93	18
Q3	1.077	1.031	1.037	1.071
	3	7	76	68
Q4	0.821	1.117	1.063	1.087
	1	1	2	150

1988-2005

1988	2005 Q1	Q2	Q3	Q4
Q1	1.269	1.249	1.191	1.156
	23	39	34	12
Q2	1.107	1.166	1.137	1.120
	3	16	45	63
Q3	-	-	1.120	1.204
	0	0	13	90
Q4	-	-	-	1.449
	0	0	0	94

1988-2010

1988	2010 Q1	Q2	Q3	Q4
Q1	1.643	1.611	1.483	1.555
	11	11	9	2
Q2	1.358	1.256	1.274	1.285
	17	18	32	16
Q3	1.170	1.088	1.189	1.151
	3	6	36	32
Q4	-	-	1.235	1.187
	0	0	1	70

Table 3.6 – continued

1998-2005

1998	2005 Q1	Q2	Q3	Q4
Q1	1.096	1.109	1.076	1.018
	42	144	84	6
Q2	0.974	1.118	1.094	1.056
	2	53	172	41
Q3	-	1.051	1.094	1.111
	0	2	73	185
Q4	-	-	-	1.288
	0	0	0	209

1998-2010

1998	2010 Q1	Q2	Q3	Q4
Q1	1.343	1.202	1.242	1.150
	43	21	16	2
Q2	1.245	1.136	1.113	1.111
	61	61	30	7
Q3	1.076	1.057	1.063	1.049
	15	44	107	36
Q4	-	0.904	1.098	1.082
	0	1	8	132

2005-2010

2005	2010 Q1	Q2	Q3	Q4
Q1	1.420	1.074	-	-
	13	2	0	0
Q2	1.137	1.087	1.074	-
	56	14	6	0
Q3	1.029	0.953	0.946	0.938
	141	166	64	2
Q4	1.000	0.929	0.916	0.844
	2	25	149	210

Table 3.7: Mean Estimates of Cost Efficiency Change by Quartile

1988-1998

1988	1998 Q1	Q2	Q3	Q4
Q1	1.023	1.003	0.974	5.457
	48	29	17	1
Q2	0.977	0.961	0.944	0.865
	21	41	93	18
Q3	0.992	1.063	0.982	0.908
	3	7	76	68
Q4	1.571	1.072	1.101	0.862
	1	1	2	150

1988-2005

1988	2005 Q1	Q2	Q3	Q4
Q1	1.018	1.051	0.971	1.057
	23	39	34	12
Q2	0.959	0.947	0.909	0.846
	3	16	45	63
Q3	-	-	0.985	0.837
	0	0	13	90
Q4	-	-	-	0.565
	0	0	0	94

1988-2010

1988	2010 Q1	Q2	Q3	Q4
Q1	1.020	0.972	1.088	2.547
	11	11	9	2
Q2	0.957	1.004	0.991	0.962
	17	18	32	16
Q3	1.016	1.138	0.989	0.901
	3	6	36	32
Q4	-	-	1.211	0.709
	0	0	1	70

Table 3.7 – continued

1998-2005

1998	2005 Q1	Q2	Q3	Q4
Q1	0.954	0.956	0.936	0.881
	42	144	84	6
Q2	0.937	0.953	0.949	0.914
	2	53	172	41
Q3	-	1.435	0.947	0.913
	0	2	73	185
Q4	-	-	-	0.721
	0	0	0	209

1998-2010

1998	2010 Q1	Q2	Q3	Q4
Q1	1.021	1.009	1.004	0.966
	43	21	16	2
Q2	1.074	1.041	0.986	1.011
	61	61	30	7
Q3	1.120	1.036	1.050	0.949
	15	44	107	36
Q4	-	1.326	1.065	0.887
	0	1	8	132

2005-2010

2005	2010 Q1	Q2	Q3	Q4
Q1	0.848	1.390	-	-
	13	2	0	0
Q2	1.045	1.001	1.151	-
	56	14	6	0
Q3	1.098	1.099	1.0828	1.180
	141	166	64	2
Q4	1.024	1.198	1.139	1.182
	2	25	149	210

Table 3.8: Mean Estimates of Cost Technology Change by Quartile

1988-1998

1988	1998 Q1	1998 Q2	1998 Q3	1998 Q4
Q1	0.929	0.915	0891	0.983
	48	29	17	1
Q2	0.912	0.924	0.916	0.902
	21	41	93	18
Q3	0.906	0.883	0.915	0.947
	3	7	76	68
Q4	0.775	1.133	0.915	0.938
	1	1	2	150

1988-2005

1988	2005 Q1	2005 Q2	2005 Q3	2005 Q4
Q1	0.719	0.758	0.753	0.742
	23	39	34	12
Q2	0.777	0.803	0.811	0.807
	3	16	45	63
Q3	-	-	0.798	0.870
	0	0	13	90
Q4	-	-	-	1.067
	0	0	0	94

1988-2010

1988	2010 Q1	2010 Q2	2010 Q3	2010 Q4
Q1	0.817	0.812	0.757	0.762
	11	11	9	2
Q2	0.739	0.723	0.719	0.685
	17	18	32	16
Q3	0.724	0.683	0.720	0.734
	3	6	36	32
Q4	-	-	0.803	0.781
	0	0	1	70

Table 3.8 – continued

1998-2005

1998	2005 Q1	Q2	Q3	Q4
Q1	0.852	0.872	0.869	0.828
	42	144	84	6
Q2	0.809	0.888	0.877	0.856
	2	53	172	41
Q3	-	0.827	0.896	0.905
	0	2	73	185
Q4	-	-	-	1.079
	0	0	0	209

1998-2010

1998	2010 Q1	Q2	Q3	Q4
Q1	0.844	0.801	0.802	0.766
	43	21	16	2
Q2	0.815	0.772	0.761	0.777
	61	61	30	7
Q3	0.762	0.764	0.766	0.776
	15	44	107	36
Q4	-	0.713	0.794	0.820
	0	1	8	132

2005-2010

2005	2010 Q1	Q2	Q3	Q4
Q1	1.208	1.042	-	-
	13	2	0	0
Q2	0.956	0.920	0.926	-
	56	14	6	0
Q3	0.895	0.862	0.859	0.842
	141	166	64	2
Q4	0.882	0.858	0.848	0.773
	2	25	149	210

Table 3.9: Revenue Efficiency Change Estimates by Quartile

1988-1998

1988	1998 Q1	Q2	Q3	Q4
Q1	0.956 48	1.066 29	1.083 17	1.981 1
Q2	1.003 21	1.077 41	1.058 93	1.011 18
Q3	0.772 3	1.038 7	1.012 76	1.011 68
Q4	1.085 1	2.396 1	1.514 2	0.906 150

1988-2005

1988	2005 Q1	Q2	Q3	Q4
Q1	1.042 23	1.137 39	1.059 34	1.057 12
Q2	1.022 3	1.070 16	1.061 45	0.996 63
Q3	- 0	- 0	1.036 13	0.968 90
Q4	- 0	- 0	- - 0	0.744 94

1988-2010

1988	2010 Q1	Q2	Q3	Q4
Q1	0.878 11	1.085 11	1.139 9	1.411 2
Q2	0.897 17	1.032 18	1.088 32	0.870 16
Q3	0.965 3	1.081 6	1.119 36	0.993 32
Q4	- 0	- 0	1.090 1	0.844 70

Table 3.9 – continued

1998-2005

1998	2005 Q1	Q2	Q3	Q4
Q1	1.149	1.115	1.062	0.911
	42	144	84	6
Q2	1.011	1.033	1.005	0.940
	2	53	172	41
Q3	-	1.204	1.016	0.960
	0	2	73	185
Q4	-	-	-	0.856
	0	0	0	209

1998-2010

1998	2010 Q1	Q2	Q3	Q4
Q1	0.947	1.082	1.038	0.835
	43	21	16	2
Q2	0.889	1.009	0.983	0.863
	61	61	30	7
Q3	0.903	1.019	1.021	0.976
	15	44	107	36
Q4	-	1.143	1.109	0.921
	0	1	8	132

2005-2010

2005	2010 Q1	Q2	Q3	Q4
Q1	0.715	0.769	-	-
	13	2	0	0
Q2	0.878	1.040	1.003	-
	56	14	6	0
Q3	0.894	1.002	1.021	0.700
	141	166	64	2
Q4	0.473	1.062	1.023	1.087
	2	25	149	210

Table 3.10: Revenue Technology Change Estimates by Quartile

1988-1998

1988	1998 Q1	Q2	Q3	Q4
Q1	1.474	1.306	1.260	2.115
	48	29	17	1
Q2	1.100	1.040	1.042	1.053
	21	41	93	18
Q3	1.181	1.059	1.045	1.063
	3	7	76	68
Q4	1.031	1.068	1.065	1.070
	1	1	2	150

1988-2005

1988	2005 Q1	Q2	Q3	Q4
Q1	1.543	1.430	1.420	1.406
	23	39	34	12
Q2	1.164	1.197	1.207	1.210
	3	16	45	63
Q3	-	-	1.245	1.249
	0	0	13	90
Q4	-	-	- -	1.423
	0	0	0	94

1988-2010

1988	2010 Q1	Q2	Q3	Q4
Q1	2.337	1.979	1.847	2.743
	11	11	9	2
Q2	1.736	1.500	1.523	1.547
	17	18	32	16
Q3	1.498	1.306	1.363	1.290
	3	6	36	32
Q4	-	-	1.316	1.272
	0	0	1	70

Table 3.10 – continued

1998-2005

1998	2005 Q1	Q2	Q3	Q4
Q1	0.998 42	1.112 144	1.121 84	1.113 6
Q2	1.042 2	1.152 53	1.169 172	1.155 41
Q3	- 0	1.150 2	1.172 73	1.162 185
Q4	- 0	- 0	- 0	1.302 209

1998-2010

1998	2010 Q1	Q2	Q3	Q4
Q1	1.542 43	1.379 21	1.421 16	1.407 2
Q2	1.435 61	1.326 61	1.327 30	1.302 7
Q3	1.245 15	1.242 44	1.220 107	1.166 36
Q4	- 0	1.147 1	1.172 8	1.189 132

2005-2010

2005	2010 Q1	Q2	Q3	Q4
Q1	1.646 13	1.289 2	- 0	- 0
Q2	1.286 56	1.188 14	1.191 6	- 0
Q3	1.154 141	1.060 166	1.035 64	1.043 2
Q4	1.048 2	1.009 25	1.005 149	0.917 210

Appendix A

Appendix: Details of Non–parametric Estimation and Inference

A.1 Dimension Reduction

Most non-parametric regression methods suffer from the well-known curse of dimensionality, a phenomenon that causes rates of convergence to become slower, and estimation error to increase dramatically, as the number of continuous right-hand side variables increases (the presence of discrete dummy variables does not affect the rate of convergence of our estimator). To help mitigate this problem, we use a dimension-reduction technique based on principal components. The idea is to trade a relatively small amount of information in the data for a reduction in dimensionality that will have a large (and favorable) impact on estimation error.

Let $J = 7$ denote the sum of the number of continuous variables on the right-hand side of Model j, excluding the ordered categorical variable T and the binary dummy variable D. For an (n x 1) vector $\boldsymbol{U}$ define the function

$$\psi_1(\boldsymbol{U}) \equiv (\boldsymbol{U} - n^{-1}\boldsymbol{U}'\boldsymbol{U})[n^{-1}\boldsymbol{U}'\boldsymbol{U} - n^{-2}\boldsymbol{U}'\boldsymbol{i}\boldsymbol{i}'\boldsymbol{U}]^{-1/2} \tag{A.1}$$

where $\boldsymbol{i}$ denotes an (n x 1) vector of 1s. The function $\psi_1(.)$ standardizes a variable by subtracting its sample mean and then dividing by its sample standard deviation. Next, let $\mathbf{A}$ be an (n x J) matrix whose columns contain $\psi_1\log((Y_1 + Y_2 + Y_3 + Y_4))$, $\psi_1\log((10 + Y_5))$, $\psi_1\log((7 \text{ x} 10^6)Y_5))$, $\psi_1\log(Z_1)$, $\psi_1\log((2 \text{ x } 10^6)Z_1)$, $\psi_1(\log(\frac{W_1}{W_3}))$, and $\psi_1(log(\frac{W_2}{W_3}))$.

Let Λ be the (J x J) matrix whose columns are the eigenvectors of the (J x J) correlation matrix whose elements are the Pearson correlation coefficients for pairs of columns of $\mathbf{A}$. Let λ_k be the eigenvalue corresponding to the kth eigenvector in the kth column of Λ, where the columns of Λ, and hence the corresponding eigenvalues, have been sorted so that $\lambda_1 \geq ... \geq \lambda_5$. Then set $\boldsymbol{P}$=$\mathbf{A}\Lambda$. The matrix $\boldsymbol{P}$ has dimensions (n x J), and its columns are the principal components of $\mathbf{A}$. Principal component vectors are orthogonal, and for each $k \in \{1, 2, ..., J\}$, the quantity

$$\phi_k = \frac{\sum_{j=1}^{k} \lambda_j}{\sum_{l=1}^{J} \lambda_l} \tag{A.2}$$

represents the proportion of the independent linear information in $\mathbf{A}$ that is contained in the first k principal components, i.e., the columns of $\boldsymbol{P}$.

Using our data, we find ϕ_k = 0.5055, 0.7220, 0.8629, 0.9247, 0.9584, 0.9862, and 1.0 for k = 1, . . . , 7 respectively. Consequently, in our non-parametric estimation of the cost function in each model, we use the first four principal components, omitting the last three. In doing so, we sacrifice a relatively small amount of informationabout 7.5 percent of the independent linear information in the samplein order to reduce the dimensionality of our estimation problem by three dimension in the space of the continuous covariates. Given the curse of dimensionality, this seems a good trade-off.[1]

A.2 Estimating Returns to Scale

Let $\boldsymbol{P}_{.k}$ denote the kth column of the principal component matrix $\boldsymbol{P}$ and define

$$\psi_0(\boldsymbol{P}_{.k}) \equiv \boldsymbol{P}_{.k}[n^{-1}\boldsymbol{P}'_{.k}\boldsymbol{P}_{.k} - n^{-2}\boldsymbol{P}'_{.k}\boldsymbol{i}\boldsymbol{i}'\boldsymbol{P}_{.k}]^{-1/2} \tag{A.3}$$

[1]The convergence rate of our local linear estimator is $n^{1/(4+l)}$, where l is the number of continuous right-hand side variables. With n = 135, 635 observations and l = 4 continuous right-hand side variables, we achieve an order of estimation error that would require 594,180 observations with five continuous right hand-side variables, 2,602,943 observations with six continuous right hand-side variables, and 11,402,792 observations with six continuous right hand-side variables.

The transformation $\psi_0(\boldsymbol{P}_{.k})$ has (constant) unit variance. Next, let z_i represent the row vector containing the ith observations on $\psi_0(\boldsymbol{P}_{.1})$, $\psi_0(\boldsymbol{P}_{.2})$, $\psi_0(\boldsymbol{P}_{.3})$, and $\psi_0(\boldsymbol{P}_{.4})$. We can now write our model as the following regression equation:

$$C_i = m(z_i, T_i, D_i 1) + \varepsilon_i \tag{A.4}$$

where the subscript i indexes observations, $C_i = \psi_1(\log(\frac{C_i}{W_{i3}}))$, and ε_i is a random error term with $\mathrm{E}(\varepsilon_i) = 0$ and $\mathrm{VAR}(\varepsilon_i) = \sigma^2(z_i)$. The function $m(z_i, T_i, D_i) = \mathrm{E}(C_i—m(z_i, T_i, D_i)$ is a conditional mean function, and can be estimated by non–parametric methods. Moreover, since the transformation from (C/W3) to C can be inverted, given an estimated value $\widehat{m}(z, T, D_1)$, straightforward algebra leads to an estimate

$$\widehat{C}(\mathbf{y},\, \mathbf{w}) = exp[\psi_1^{-1}(\widehat{m}(z, T, D_1)]. \tag{A.5}$$

To estimate returns to scale, we need merely estimate the measure $S(\theta \mid \mathbf{y}_0, \mathbf{w}_0)$ defined earlier by replacing $C(\mathbf{y}_0, \mathbf{w}_0)$ and $C(\theta\mathbf{y}_0, \mathbf{w}_0)$ on the right-hand side of (2.4) with estimates $\widehat{C}(\mathbf{y}_0, \mathbf{w}_0)$ and $\widehat{C}(\theta\mathbf{y}_0, \mathbf{w}_0)$ obtained from (A.5).

In order to estimate the conditional mean function in (A.4), suppose (for the moment) that the time variable T and the binary dummy variables D_1, D_2 do not influence the value of the conditional mean function $m(z_i, T_i, D_1, D_2)$, so that we can write the conditional mean function on the right-hand side of (A.4) as $m(z)$. Both the Nadaraya-Watson (Nadaraya, 1964; Watson, 1964) kernel estimator and the local linear estimator are special cases of local polynomial estimators; with the local linear estimator, the local polynomial is of order 1, while with the Nadaraya-Watson estimator the local polynomial is of order 0. The local linear estimator has less asymptotic bias, but the same asymptotic variance, as the Nadaraya- Watson estimator.

To illustrate the local-linear estimator, momentarily ignore the discrete covariates in (A.4) and write the conditional mean function as $m_*(z)$ in a neighborhood of an arbitrary point z_0:

$$m_*(z) \approx m_*(z_0) + \frac{\partial m_*(z)}{\partial z}(z - z_0) \tag{A.6}$$

This suggests estimating the conditional mean function at z_0 by solving the locally weighted least squares regression problem

$$[\widehat{\alpha}_0 \ \widehat{\alpha}]' = argmin \sum_{i=1}^{n} [C_i - \alpha_0 - (z_i - z_0)\alpha]^2 K(|\mathbf{H}|^{-1}(z_i - z_0)) \tag{A.7}$$

where $K(\cdot)$ is a piece-wise continuous multivariate kernel function satisfying $\int ... \int_{\mathbb{R}^l} K(u)du = 1$ and $K(u) = K(-u)$, $u \in \mathbb{R}^l$; $\mathbf{H}$ is an l x l matrix of bandwidths; α_0 is a scalar, and α is an l-vector.

The solution to the least squares problem in (A.7) is

$$[\widehat{\alpha}_0 \ \widehat{\alpha}]' = (\mathbf{Z}'\Phi\mathbf{Z})^{-1}\mathbf{Z}'\Phi\mathbf{C}, \tag{A.8}$$

where $\mathbf{C} = [C_1 ... C_n]'$, $\Phi = \text{diag}\ [K(|\mathbf{H}|^{-1}(z_i - z_0))]$, and $\mathbf{Z}$ is an n x $(l+1)$ matrix with ith row given by $[1\ (z_i - z_0)]$. The fitted value $\widehat{\alpha_0}$ provides an estimate $\widehat{m_*}(z_0)$ of the conditional mean function $m_*(z_0)$ at an arbitrary point z_0.[2]

[2]The fitted values in $\widehat{\alpha}$ provide estimates of elements of the vector $\partial m(z_0)/\partial z$. However, if the object is to estimate first derivatives, mean-square error of the estimates can be reduced by locally fitting a quadratic rather than a linear expression (see Fan and Gijbels, 1996 for discussion); this increases computational costs, which are already substantial for the local linear fit. Moreover, determining the optimal bandwidths becomes more difficult and computationally more burdensome for estimation of derivatives. See Hardle (1990, pp. 160–162) for discussion of some of the issues that are involved with bandwidth selection for derivative estimation.