NORTHERN ILLINOIS UNIVERSITY

THE ROLE OF SCOPE AND SCALE ECONOMIES IN RECENT LARGE BANK MERGERS

A Thesis submitted to the

University Honors Program

In Partial Fulfillment of the

Requirements of the Baccalaureate Degree

With University (Upper Division) Honors

Department of Finance

by Daniel Eichelberger

DeKalb, Illinois

May 9, 1992
ABSTRACT

The recent mega-merger activity in the U.S. banking industry raises many issues. Most important is the question of whether these mergers result in more profitable banks. A review of the literature on cost savings due to economies of scope and scale suggests that only those savings from the diversification of risk are present. The savings due to this diversification are substantial, but we also need to look at other areas of cost savings. Theories such as the information hypothesis, the market-power hypothesis, the inefficient-management hypothesis, and the too big to fail theory lead us to believe that the merging of these banks can substantially reduce costs. If the recent mega banks prove to be profitable, we may see the average size of banks increase. This may also lead to the dominance of a few super-banks, those that have already begun this trend.
I. INTRODUCTION

The issue of whether economies of scope and scale exist in commercial banking has long been debated. If these economies do in fact exist, it offers one possible explanation for the rash of recent mega-mergers the commercial banking industry has seen. For example, Chemical and Manufacturers Hanover banks merged, resulting in a $137 billion bank. NCNB and C&S Sovran also merged, forming Nation's Bank, with $118 billion in assets. Additionally, there was the marriage of Bank America and Security Pacific, combining assets of over $196 billion resulting in the second largest U.S. bank (Matthews 1991). If, however, economies of scale and scope are found not to exist for these mega banks, alternative explanations for these mega-mergers should be investigated.

This paper focuses on the economy of scale and scope issue by first reviewing some of the more widely-accepted studies on the subject. A conclusion based on this literature will then be provided. Finally, other explanations for the recent mega-mergers will be discussed. These include profits from information, market power, tax benefits, the elimination of inefficient managers, and insurance subsidies.
II. LITERATURE REVIEW

To aide the reader in understanding these studies, some basic terminology will now be presented. Economies of scale are present when a firm's average cost of production declines as the quantity of its output increases. Economies of scope are said to exist when the production of two or more services together costs less than producing them separately. Studies in this area commonly measure cost savings using cost functions (also called production functions), which represent an attempt to measure costs of inputs and outputs in various amounts and combinations. If the cost functions reveal lower costs due to increased outputs or by combining outputs, economies of scale or scope are said to exist.

Some of the cost measures use single product functions while others use multi-product functions. Single product measures estimate all output using one figure such as total assets. Multi-product production functions use several figures to measure output, such as deposits, transactions, and overhead. A traditional U-shaped cost curve describes the costs of producing a unit of output. At small levels of production, each unit of production costs more because overhead are spread over fewer units. This represents the left side of the U curve. Unit costs decrease as production increases, moving down the curve. Once economies of scale have been exhausted, unit costs begin to rise due to inefficiencies. This represents the right side of the U curve.
Clark (1984) performed a study on the estimation of economies of scale in the banking industry. Clark felt that the traditional Cobb-Douglas production function lacked credibility because it facilitated the estimation of the output elasticity of costs. There are no sufficient reasons to support this estimation. Clark, therefore, developed a Box-Cox generalized functional form methodology to test the validity of the Cobb-Douglas production function.

The article refers to the difficulties in determining the appropriate definition of bank output. Three approaches have been used in the past. Many early studies used unweighted measures, such as total assets, in estimating output. The problem with this measure is that it ignores the multi-product nature of commercial bank output. A second approach is to assign each bank service with a separate production function. This approach captures the multi-product aspect of output but ignores cost savings related to joint production of services. The third approach constructs a weighted index of bank output using data from the income statement and balance sheet. The index represents total revenue from earning assets adjusted for price differences among banks. This method solves the problems of the previous two approaches but fails to include the nonlending activities that make up to 10% of a banks income. Clark adjusts for this by including the income from nonlending activities to the weighted index.
The data used in this study were taken from the Report of Income and the Report of Condition published by the Board of Governors of the Federal Reserve System, 1972-1977. The banks in the sample included 1,205 unit banks located in 57 standard metropolitan statistical areas. Bank sizes ranged from $7 million in total assets to $425 million.

This study used all three approaches to measuring output in testing the Cobb-Douglas production function and the Box-Cox functional form. Three conclusions were drawn. First, both the Cobb-Douglas and Box-Cox functions revealed similar results. Therefore, the limitations of the Cobb-Douglas production function do not discredit its findings. Second, the cost estimates were highly insensitive to the choice of bank output used. This suggests that cost savings are not very significant in nonlending activities or in joint production of services. Finally, both cost functions reveal small economies of scale in the banking industry.

Benston, Hanweck, and Humphrey (1982) examined bank scale and scope economies using a translog cost function. This measure was hoped to overcome the limitations of previous studies. Benston et al. suggest that prior studies were limited in three ways. First, these studies did not measure total banking costs. Only those costs of specified activities were used. Second, branch banks were combined with unit banks. Finally, past cost equations did not allow for a U-shaped curve.
The translog cost equation permits the estimation of a U-shaped cost curve, total banking costs, and allows economies of scope and scale to vary by bank size. This cost function also separates economies of scale and scope between branches and unit banks. The authors, therefore, believe the translog function to be a more comprehensive measure.

The data used in the study was taken from the Federal Reserve's Functional Cost Analysis (FCA) program. The sample of banks were arranged in nine deposit-size classes. Banks over $100 billion in deposits were eliminated from the study due to lack of representation in the FCA data.

The cost variable used was total operating costs, measured as all operating expenses other than interest payments. Operating costs include the costs associated with demand deposits, time deposits, real estate loans, and commercial and industrial loans. These costs account for 72% of total operating expenses, but are subject to some biases. First, historical costs are used in estimating assets. Second, opportunity costs of services performed by corresponding banks are omitted. Finally, the cost of risks are not directly measured.

Commercial bank output was measured using a Divisia Multilateral Statistical Index. This index is the sum of the bank's accounts corrected for the difference in costs for deposits and loan accounts among banks. The Divisia index corrects the problem of assuming that one dollar of demand deposits has the same costs as one dollar of time deposits or loans.
The results of this study show that economies of scale and scope are different for branch and unit banks. Branches were found to experience small economies of scale and scope. Unit banks, however, were found to experience diseconomies of scale at banks with more than $50 million in deposits. Unit banks experienced no economies of scope in their operations. When many branches were brought together, diseconomies similar to those of unit banks occur.

Gilligan, Smirlock, and Marshall (1984) suggest that previous studies lacked proper cost functions to measure the multi-product nature of the banking industry. Their study therefore tried to incorporate the multi-product cost into their evaluation.

In developing a cost measure, the authors looked to Baumol's concept of Ray Average Costs (RAC). This formula groups average costs to the multi-product bank without affecting total output indices. "RAC relates total production costs to a proportionate increase in all output." This is different from the current multi-product production function which generalizes all scale economies. RAC also separates lower costs due to cost complementarities. Cost complementarities are the cost savings due to producing one or more products together. These cost savings are called economies of scope.

Gilleigan et al. used the RAC measure because of problems with the popular multi-product cost functions. The popular multi-product cost function was developed by Benston (1965) and
Bell and Murphy (1968). This function measures costs as a group of individual Cobb-Douglas functions. Each bank activity is given a separate cost function. Unfortunately, lower costs due to scope economies are present in the measure. The measure has also been criticized for not measuring total costs.

Gilligan et al. used specific data from the Federal Reserve's Functional Cost Analysis (FCA) program including a sample size of 714 banks with deposits under $1 billion. In measuring output, the authors used an average account size variable which is measured as the dollar value of deposits divided by the number of deposit accounts. An appropriate measure of cost was needed so a sum of non-interest expenses allocated to deposits and loan activities was used. Finally, the price of capital was approximated by the rental cost of bank and office buildings in nine geographic regions.

Using the RAC data, it was found that economies of scale do exist at unit banks and branches with deposits between $25 million and $100 million. The data also reveal diseconomies of scale exist for banks with deposits greater than $25 million. These findings are consistent with the traditional U-shaped cost curves.

Berger, Hanwick, and Humphrey (1987) identified the lack of a composite cost measure in determining whether economies of scale and scope exist. The study developed two new cost measures that can better determine the cost efficiency of banks that vary in scale and product mix simultaneously.
In developing a proper cost study, banks with any combination of scale and product mix must be included in the cost measure. Past studies have either used Ray Scale Economies or Scope Economies. Ray Scale Economies compare the costs of firms that differ in scale but not product mix, while Scope Economies compare the costs of firms that differ in product mix but have the same scale for each output. Therefore, these prior studies have only been able to compare costs of banks with exactly the same combinations of output.

The authors used scale, scope, and product mix measures that incorporate various output categories, solving the problems addressed. Data on nine different sizes of banks and the overall average were used. A six equation model using cost accounting data also attempts to reduce the estimation error. Finally, "dollars intermediated was used as an output measure without the usual constraint that large and small accounts cost the same per dollar."

A Ray Scale Economies measure was used in the study to compare with the two new measures. This measure is meant to be used as a control variable in the study. A second cost measure used was the Expansion Path Scale Economies. This new method measures the expansion or growth path of a bank's output and product mix as size increases. The key is the isolation of incremental costs corresponding to incremental output. A third cost measure used was the Expansion Path Subadditivity Model. This measure gives the proportional cost increase from two firms
producing the same product mix as one firm. All three measures use two ways of analyzing outputs and costs (Production and Intermediation), and at two levels of the firm (Plant and Firm), and at two competitive environments (Branching and Unit Banks).

Under the production approach, banks produce accounts of various sizes by processing deposits and loans, incurring capital and labor costs. Operating costs are specified in the cost function and numbers of accounts are used as the output measure, while average account sizes are specified to control for other account characteristics.

The intermediation approach has banks intermediate deposited and purchased funds into loans and other assets. Total operating and interest expenses are used as costs, and the amount of dollars is used as the output measure. The intermediation approach is preferred in a competitive environment because it includes both operating and interest costs.

The data used in all measures in this study have been taken from the Functional Cost Analysis (FCA) data provided by the Federal Reserve. The FCA data is from 1983 and contains statistics on 413 branches and 214 unit banks. The largest banks are not included in the data.

Results for Ray Scale Economies are as follows: Under the Production approach, banks show economies of scale only at the small bank level. Small banks are considered as banks with less than $50 million in deposits. The results are the same at the firm level. The intermediation approach reveals similar results
but show diseconomies of scale as banks grow. Unit banks, however, show significant scale diseconomies for large banks. Large banks are those with over $100 million in deposits.

Expansion Path Scale Economies show erratic results. Banks alternate between economies and diseconomies of scale as product mix changes with size. The data shows that incremental costs are inconsistent with incremental product mixes and size. This measure found insignificant economies of scope at small banks and diseconomies of scope at large banks.

Expansion Path Subadditivity measures show diseconomies of scale. Using the intermediation approach, slight diseconomies of scale were found at banks with deposits above $50 million. The production approach, however, revealed larger diseconomies of scale at the same level.

In summary, all of the results are generally consistent with showing insignificant economies of scale at small banks and diseconomies of scale at large banks. No scope economies were found at small banks with scope diseconomies occurring at large banks. The study did reveal unrealistic diseconomies of scope and scale at the largest banks. This can be attributed to the difficulty in extrapolating the results to larger banks.

Much of the data used in these studies has been outdated. Most of the past studies use data from the 1970's or earlier. With the rapid changes in the banking industry, it is difficult to draw conclusions based on past data. In trying to predict economies of scale and scope in recent mergers, it is beneficial
to perform tests using more recent information. Buono and Eakin (1989) try to solve this problem and others in their analysis.

Buono and Eakin try to overcome the shortcomings of past studies in developing their cost function. They suggest the following problems exist with the past studies. Clark (1984) ignores the multi-product nature of commercial banking. Clark showed a high correlation among bank output measures. Therefore, he felt that a single output measure could be used. This becomes a problem due to the fact that banking activities are produced in various amounts. Benston et al. (1982) used their Divisia index to construct an aggregate measure to handle this multi-product nature in banking. The affect on cost of varying the mix of outputs cannot be measured using an aggregate index. Gilligan et al. (1984) use a translog cost function to measure multi-product costs, but this measure lacks the flexibility to handle economies of scope. Berger et al. (1987) estimate the overhead and direct cost equation for each output which requires the previous allocation of costs associated with the previous product mix. This measure lacks the ability to handle product variation.

Buono and Eakin feel their dual cost function can solve these past problems by incorporating more recent data and by modeling banks as a three-product firm.

In estimating their cost function, Buono and Eakin use 1985 Functional Cost Analysis (FCA) data of 613 member banks of the Federal Reserve System. This includes 387 branch banks and 226 unit banks. The banks are modeled as a three product firm.
Outputs are loans, investments, and transaction deposits. Loans and investments are then weighted by their percentage of revenue. Not enough data is available for the weighting of transaction deposits.

Buono and Eakin post results consistent with many prior studies. The dual cost function reveals diseconomies of scale and scope at the unit bank level. This is inconsistent with the results for the branch level where economies of scale and scope exist. This suggests that banks should either increase output at existing branches or decrease the number of branches.

Srinivasan and Wall (1992) try to conduct a study of economies of scale in large banks which have merged. The authors claim that recent mergers used cost savings as one of the reasons for merging. The article, therefore, tried to measure the non-interest expenses of merging banks before and after the merger to determine if there were cost savings.

The data used in the study were taken from the year-end Reports of Income and Condition (Call Reports) compiled by the FDIC. Only commercial bank and bank holding company mergers occurring between 1982 and 1986 were used. Several criteria had to be met to satisfy the final sample. First, both merging banks had to exceed $100 million in total assets. Second, both banks had to be previously unaffiliated. Finally, only domestic mergers were allowed in the sample. The ratio of non-interest expense to total assets is used to measure the cost savings from bank mergers. Non-interest expense is the sum of salaries,
premises, and other expenses. The change in non-interest expense is examined starting two years prior to the merger and ending four years after. The year of the merger is not included due to complications in gathering correct figures in that year.

The findings of the study show that the non-interest expense ratio for merging banks increased after the merger occurred. The results also suggest that the acquiring banks were more efficient than the industry prior to the merger but are no more efficient afterwards. The study, therefore, states that economies of scale do not exist as a result of mergers. The authors do admit to two limitations. First, the sample does not include any mergers comparable to the size of recent mergers. Second, the merging banks may not have intended on reducing costs. It is also important to note that there are many more expenses to be considered than just those mentioned in the study.

Shaffer and David (1991) also perform a study on economies of scale and scope in large commercial banks. They, however, do not limit the data to those banks that have merged. The authors look to the 100 largest U.S. banks in testing their cost measures. The results of the study differ dramatically from those of Srinivasan and Wall (1992).

Many prior studies have relied on Functional Cost Analysis (FCA) data. This could be a reason for the common view that larger banks experience diseconomies of scale. The FCA data involves almost exclusively small banks. Banks with more than $1 billion in assets are not present in the data. The average size
bank in the FCA program has approximately $100 million in assets. Shaffer and David use Call Report data compiled by the FDIC. Specifically, the 100 largest commercial banks in 1984 were used. The banks range in size from $2.5 billion to $120.6 billion in assets.

The authors use a scalar output measure to capture bank production. This means that a single figure is used to represent all various output figures. Total assets are used to represent all banking output. The rationale for this approach is that total assets is easy to track, and past studies have used several output measures to test the scalar output approach finding it to be an acceptable measure. Total deposits are used to represent the input variable. Deposits are found to be positively correlated with bank expenses. This represents the risk differential paid on deposits greater that the FDIC insured $100,000 limit. Theoretically, larger banks should be able to attain lower risk through diversification. The reduced risk should therefore result in lower cost of uninsured deposits.

The results are extremely different from past studies that show economies of scale exhausted above $100 million in assets. Allowing for the market to price the risk of uninsured deposit rates has revealed a previously unmeasured benefit of size. This is due to the fact that larger banks can diversify their assets and become less risky than smaller banks. The authors have found scale economies to exist at banks with assets between $15 billion and $37 billion. The historic viability of large banks coupled with recent merger activity seem to support these findings.
III. SUMMARY OF LITERATURE

The review of the literature provided shows results to be mixed. Each study used different variations of cost functions, inputs, and outputs to test for economies of scale and scope. Clark (1984) found no significant economies of scale or scope. Benston et al. (1982) concluded that there were no economies of scale, and diseconomies of scale at banks with more than $25 million in assets. Gilligan et al. (1984) found economies of scale between $10 million and $50 million with diseconomies above $50 million in assets. Berger et al. (1987) found only diseconomies of scope and scale above $100 million in assets. Buono and Eakin (1990) found diseconomies to exist at banks, but scope and scale economies at their branches. Srinivasan and Wall (1992) found no economies of scope or scale at large banks that have merged. Shaffer and David (1991) tested only the largest banks and included cost savings due to diversification in their study. They found economies of scale at banks with up to $37 billion in assets.

The studies can be grouped into two categories. The first category includes those studies that show diseconomies or no economies of scale or scope for large banks. This includes all the studies except Shaffer and David (1991). The second category would be the study that finds significant economies of scale at large banks. The next step is to isolate the relevant differences between the categories. The only answer to the measurable difference between the categories is that Shaffer and David (1991) include cost savings from diversification. All
other aspects result in only minor differences. For example, using a single output measure rather than several to estimate banking costs is immaterial.

Cost savings due to diversification is a very important variable. Therefore, I believe that Shaffer and David (1991) are correct in estimating economies of scale for banks with up to $37 billion in assets. This, however, does not explain the mega-mergers mentioned in the introduction. Bank America and Security Pacific have assets of $196 billion, well above the $37 billion. This would suggest that there are other areas of cost savings to encourage these recent mega-mergers.
IV. ALTERNATIVE HYPOTHESES FOR MEGA BANK MERGERS

Economies of scale and scope are not the only reasons for the recent merger activity. There are several other motives for merging. The following is an explanation of several hypotheses.

Information Hypothesis

The information hypothesis states that the shares of some banks are incorrectly valued by the market because the public does not possess valuable private information (Hawawini 1990). This suggests that the stock market operates in a semi-strong efficiency. A market that is semi-strong form efficient is one where stock prices reflect all publicly available information but not information that is held private. If the acquiring bank is aware of private information that leads them to believe the target bank is undervalued, a situation exists where the merging of the banks results in higher profits.

Market-Power Hypothesis

The market-power hypothesis is a theory that horizontal mergers create a monopolistic concentration of power by reducing the number of competing banks in the industry (Hawawini 1990). As competition decreases, the merged banks can raise their prices and lower their costs. This theory only seems to hold true in rural areas where banks are less concentrated.
Tax Hypothesis

The tax hypothesis suggest that banks merge to reduce their tax base (Hawawini 1990). If one bank experiences high profits while the other bank experiences a loss, the tax base is reduced by the merger. This, however, does not explain the merging of two profitable banks.

Inefficient-Management Hypothesis

The inefficient-management hypothesis states that bank mergers will result in the elimination of poor management (Hawawini 1990). The theory is that the managers of duplicating jobs will be fired due to the merging of two banks. This allows the merged bank to choose the most efficient managers between the old banks. This theory seems practical as we see many middle managers being fired as a result of the recent mergers. It has yet to be seen, however, whether the inefficient managers were the ones being fired.

Too Big to Fail

A bank that is so large that its failure would disrupt the nation's economy is considered too big to fail (Boyd 1991). The government will not let these banks fail and infuses public money into the bank to keep it solvent. The government, by their actions, in essence insures all deposits above the $100,000 FDIC limit. The bank, however, only pays for insurance up to the $100,000 limit. The added risk is subsidized by the government. The public identifies these too big to fail banks and invests...
their money because they know their deposits are fully insured. The added insurance coverage and increased deposits give the too big to fail banks a significant cost saving advantage over smaller banks.
V. CONCLUSION

The recent trend of mega-mergers between U.S. banks suggests that there are benefits to increased size. If mega-banks are more profitable, we may see the U.S. banking industry dominated by a few super-banks. In reviewing the literature on economies of scope and scale as a possible explanation for recent large mergers, only the cost savings due to diversification of risk are found to exist. The ability of mega-banks to attract deposits while paying a lower risk premium results in substantial cost savings. Additionally, mega-bank mergers may also receive other benefits besides those of scope and scale economies. The merging banks can increase profits if one bank is incorrectly priced by the market. The mega banks can enjoy more market concentration, resulting in a monopolistic environment. Inefficient managers are eliminated through the duplication of jobs. Finally, the mega banks can be considered too big to fail and receive an insurance subsidy by the government in the form of reduced costs of acquiring funds. These savings may result in more mega-mergers and an increased market share for these leading banks.
BIBLIOGRAPHY


