Efficiency Analysis of Financial Institutions: A Review of Empirical Studies

Allen M. Featherstone

Proceedings of a Seminar sponsored by
North Central Regional Project NC-207
"Regulatory, Efficiency and Management Issues Affecting Rural Financial Markets"
Minneapolis/St.Paul, MN
September 26-29, 1992

Food and Resource Economics Department
Institute of Food and Agricultural Sciences
University of Florida

September 1993

Efficiency Analysis of Financial Institutions: A Review of Empirical Studies

Allen M. Featherstone*

The optimal size and scale of financial institutions has received considerable attention in the finance literature. Given the recent wave in financial institution consolidation, there is much interest in optimal firm size. The major concern rests in whether consolidation is moving firms down their cost curves to achieve greater efficiency, or whether consolidation is resulting in greater market power without achieving cost savings. The study of the production technology of financial institutions can determine whether and to what degree economies of size exist.

Generally, empirical studies have used either duality theory with the estimation of cost functions or nonparametric estimation methods to assess efficiency in the financial services industry. The purpose of this study is to review the banking literature to grasp what is known about scale economies. This study will focus on the estimation of indirect multi-product cost functions using duality.

Clark (1988) reviewed 13 studies that have attempted to measure economies of scope for commercial banks, credit unions, and savings and loan associations. Clark found that these studies offered four broad conclusions: 1) overall economies of scale exist at low levels of input, 2) no consistent evidence of economies of scope, 3) some evidence of cost complementarities, and 4) the results seem to be robust among financial institutions.

Humphery also reviews studies which examine the issue of bank economies of scale. Humphery finds that little cost savings exist for increases in size alone. He finds that significant benefits accrue from loan diversification. Humphery also finds that the differences in cost structure within the same size category is large compared to measured cost economies.

The paper will be organized in the following manner. First, multiproduct cost concepts will be briefly discussed. A discussion will follow of the data and procedures used in the estimation of indirect multi-product cost curves. The paper will summarize some of the empirical findings from the banking literature. Finally, the paper will conclude with an assessment of the strengths and weaknesses of these studies and provide comments on future research needs for those interested in agricultural banks.

Multiproduct Cost Concepts

Multiproduct cost concepts did not arise until the early 1980s (Baumol, Panzar, and Willig). Multiproduct cost concepts were able to develop only after the development and application of

---

*Allen Featherstone is an associate professor in Department of Agricultural Economics, Kansas State University. The author would like to acknowledge the library work of Crystal Rempe.

duality theory. Rigorous treatment of duality originated in 1953 with a book by Ronald Shephard. This allowed a rapid expansion of the classes of functional forms available for empirical estimation of production relationships. The class of flexible functional forms, which are based on 2nd-order Taylor series approximations, include the translog, generalized Leontief, and the quadratic. The translog is the most commonly applied functional form used in multiproduct cost analysis of the banking industry.

In a multiproduct framework, economies of scale can arise from two sources: product-specific economies and/or economies of scope. Product specific economies are present if the per unit cost of producing an output declines as the output increases. In a multiproduct framework, product specific economies are measured by defining what is known as incremental cost. The incremental cost for the ith output (IC_i) is defined as the cost of producing the entire multiproduct output bundle (C(Y)) minus the cost of producing all of the outputs except the ith output ($C(Y_{N-i})$). Formally:

$$1) \qquad IC_i = C(Y) - C(Y_{N-i}) \text{ where } Y_{N-i} = (Y_1,\ldots,Y_{i-1},0,Y_{i+1},\ldots,Y_N).$$

Product-specific economies of scale (S_i) are then determined by taking the average incremental cost of producing the ith output (IC_i/Y_i) divided by the marginal incremental cost of producing the ith output ($\partial C/\partial Y_i$). Formally:

$$2) \qquad S_i = \frac{(IC_i/Y_i)}{(\partial C/\partial Y_i)}.$$

If S_i is greater than one than product specific economies of scale are said to exist. Product specific economies of scale are most analogous to the single output case of scale economies. This measure can be expanded to include subsets of products if desired.

Economies of scope (diversification) arise from cost savings obtained from the simultaneous production of several outputs together. Economies of scope ($SC_i(Y)$) exist if the cost of producing the optimal level of outputs in "individual firms" is greater than the cost of producing the same optimal output levels in a multiproduct firm. Formally for a two product firm, if

$$3) \qquad C(Y_1) + C(Y_2) > C(Y)$$

then economies of scope exist, where $C(Y_1)$ is the cost of producing output 1 in a single product firm and $C(Y_2)$ is the cost of producing output 2 in a single product firm. For 2 outputs, economies of scope ($SC_N(Y)$) are defined as:

$$4) \qquad SC_N(Y) = [\, C(Y_1) + C(Y_2) - C(Y) \,]/C(Y).$$

If $SC_N(Y)$ is greater than zero than economies of scope are said to exist. This indicates the relative increase in cost from a splintering of production into separate groups or the relative cost savings of multiproduct production.

Both economies of scope ($SC_N(Y)$) and product-specific economies (S_i) can be combined to give an overall measure of the returns to scale for an individual firm. These are also referred to as scale economies (S_N). Formally, the measure of economies of scale for a two output firm is:

$$5) \qquad S_N(Y) = \frac{\alpha_1 S_1(Y) + (1-\alpha_1) S_2(Y)}{1 - SC_N(Y)}$$

where α_1 is the first firm's output time the marginal cost as a proportion of the sum of all outputs multiplied individually by their marginal cost.

Economies of scale can then arise under multiple scenarios. If economies of scope are equal to zero, then economies of scale will exist if one of the outputs has constant returns to scale and the other output has increasing returns to scale. Economies of scale can also arise if both outputs have constant returns to scale if economies of scope exist.

Data and Procedures

The data and procedures may limit the usefulness of findings from studies measuring economies of scope and scale in the banking literature. Understanding the differences in data used and procedures may help to explain differences in results obtained by various authors.

Data

The definition of outputs and costs, the treatment of branch banking, and data sources vary widely across studies of economies of scale in the banking industry. Two approaches exist for defining inputs and outputs: 1) the production approach and 2) the intermediation approach. The production approach measures output by the number of various accounts and considers only operating cost. The intermediation approach measures output as the size of accounts (dollars) and measures costs as both operating and interest cost. The type of data and the method of approach are inter-related. Clark (1988) suggests that conceptually the approaches are very different. He suggests that the choice of approach may be one of the reasons for differences in empirical results. However, Benston, Hanweck, and Humphrey and Berger, Hanweck, and Humphrey found that similar scale economy results were obtained using both approaches.

A majority of the studies examined use one of two types of data: 1) Call Report (Income and Conditions Reports for Banks) and/or financial statement data or 2) Functional Cost Analysis (FCA) data. The FCA data is constructed using simple cost accounting techniques to allocate expenses to several banking functions (Clark 1988). Clark suggests that generalizing results from FCA data may be inappropriate because of self selection bias in the participation of subscribing banks, and the problem of introducing bias into parameter estimates due to imprecise cost allocation methods. Smaller banks tend to participate in the FCA.

Clark (1988) suggests that the use of Call Report data offers a wider range of bank size. In addition, uniform reporting requirements are imposed. However, Clark notes that the Call

Report data are not without problems. Data is not available on the number of deposit or loan accounts. Clark suggests that account size and bank size are correlated. Thus, economies of scale are likely overstated. Secondly, data on banking functions such as safety deposit box and trust activity have only recently begun to be reported. Thus, some of the banking functions are not able to be represented as outputs. A final problem with the Call Report data and with any cost function study using accounting data is the development of input prices. It is somewhat questionable whether meaningful proxies can be developed for price given aggregation of inputs.

A 1987 multiproduct study of the California savings and loan industry is used to illustrate the methods used to construct data (Mester). This study defines a three output and four input firm. The outputs are mortgage loans, other loans, and cash, securities, and real estate investments in excess of the minimum liquidity requirements. These outputs are measured in average dollars at the end of 1982, thus an intermediation approach was used in this study. The inputs are labor, capital, demand deposits (passbook and NOW accounts), and term deposits.

The prices for the inputs were determined as follows. For labor, the wage was determined as the weighted average of the county wage in which the S&L had a branch. A crude price of capital was determined by adding rent, depreciation, utilities, equipment, and furniture expenditures divided by average dollar deposits during 1982. Mester notes that a better approximation would be the average rent per square foot of office space, however, this data was not available. The measure of the price of demand deposits was determined by subtracting service charges paid on NOW accounts from total interest and commissions and dividing by average demand deposits. The price of term deposits was determined by dividing total interest by average term deposits for the year. A growth variable was also included in the analysis to account for changes in the size of the institution. This variable was calculated by taking the log of 1982 assets divided by 1981 assets. A branching variable was also included to examine the effect of the number of branches on cost.

The study by Mester illustrates the difficulty of defining the price variables for the inputs. This study is fairly representative of the literature and the methods used to define the variables. In studies of agricultural banking, defining prices will also be a difficult issue.

Procedures

Closely related to the problem with the determination of the prices of the inputs is the level of aggregation used in defining both inputs and outputs. Ideally, every output and input that a financial institution uses or produces would be treated on a disaggregated basis. The number of parameters estimated increases at an increasing rate as additional inputs or outputs are defined. Thus, the number of observations available require the number of inputs and outputs to be limited.

Another problem commonly encountered in the studies of financial institutions is the use of the translog cost function. Outputs are logged in the estimation process. If a financial institution does not produce an output, the log of that output quantity (zero) is undefined. This problem

also arises when calculating incremental costs. To get around this problem, the commonly accepted technique is to replace zero outputs with a sufficiently small nonzero value. Cowing and Holtmann, Akridge and Hertel, and Schroeder set zero output values equal to 10 percent of the geometric mean. A drawback of this procedure is that bias may be introduced into the parameter estimates.

A final problem that is common in multiproduct cost function studies is the significance of individual parameter estimates. Clark reports that multicollinearity is often a problem in the estimation. In a system requiring 36 parameter estimates only four of the estimates were significant at the 5% significance level. Authors examining other industries also find low levels of significance in parameter estimates. At the 10 percent level of significance, Cowing and Holtmann, Akridge and Hertel, and Deller et al. find that only 30%, 35%, and 59% of the estimates to be significant. Bootstrapping techniques offer the opportunity to allow confidence intervals to be placed around estimates of economies of scope and product specific economies (Efron, Eakin et al., Schroeder).

A final class of problems with multiproduct cost functions are the use of flexible functional forms and the use of econometrics. A Taylor's series approximation suggests that the error of the remainder increases as it moves away from the point of approximation. Using, regression to estimate a Taylor's series tends to make the errors of approximation equal over the entire data range (White). This problem can be alleviated using the Fourier flexible form (Gallant).

A related problem also arises due to the use of econometrics. Usual econometric procedures result with errors being both above and below the isoquant. However, when estimating the cost function and related isoquants, the most efficient or frontier function is appropriate. Thus, errors should only be above the cost isoquant. Using methods which estimate a frontier function allows firm inefficiency to be measured using Farrell's approach (Akridge).

Results from Banking Studies

Clark (1988) provides an excellent discussion of thirteen financial institution studies which examine economies of scale and scope. Clark qualitatively examines what these studies found. This review provides an updated analysis of more recent studies and will provide more detail on quantitative measures of economies of scope and scale. The discussion of results from empirical banking studies will be broken down in the following manner. First, other studies that do not use the multiproduct cost methodology but provide insight into the structure of financial institution production technology will be discussed. Following, the five studies which estimate the measures of economies of scale, economies of scope, and product specific economies of scale are discussed.

Other Studies

Clark (1984) examined the sensitivity of output elasticities to cost under various measures for a single output and different functional forms using Call Report data. Clark found output

elasticities to be fairly robust to functional form and output measure choices. A study by Kilbride, McDonald, and Miller extend Clark's work by examining the sensitivity of results to econometric procedures, examining a different data period, and considering differences based upon the type of bank organization. They find that scale economies are sensitive to bank organization.

Nelson also examines whether branch variables are important determinants in the bank cost function. Nelson's work supports that of Kilbride, McDonald, and Miller by finding substantial economies of scale at the branch level, but no economies from expansion by branching.

Goldstein, McNulty, and Verbrugge examine economies of scale in the savings and loan industry using a single measure of output (total assets). They find scale economies for small institutions which persist across medium and larger institutions.

Gilligan and Smirlock (GS) examine economies of scale in commercial banking in unit bank states in a multiproduct framework with two outputs: loans outstanding and total dollar amount of securities held. No inputs or input prices enter the cost function estimation. GS did find global economies of scope suggesting the studies which use a measure of one output or an index of one output are not appropriate. Also the allocation of costs in functional cost analysis may be suspect due to non-proportional outputs. Scale economies are found in small banks while diseconomies of scale are found in larger banks.

Shaffer and David (SD) estimate economies of scale for the 100 largest U.S. commercial banks. They use hedonic terms to correct for the aggregate nature of Call Report data. SD use bank assets as the measure of output and labor, and capital as inputs. They include qualitative variables such as the uninsured deposits to insured deposits ratio, state banking laws, relative volume of problem loans, and asset mix to adjust for differences in banks not accounted for in the inputs and output defined. SD find the minimum efficient scale is between $15 and $37 billion in assets. These findings were robust to the various hedonic variables examined.

Dowling and Philippatos examine economies of scale in the U.S. savings and loan industry. They defined output as total assets. They found that economies exist for S&Ls, the minimum efficient scale decreased from 1979 to 1983, and substantial changes have occurred in operating costs.

Multiproduct Cost Studies

Studies by Mester; Murray and White; Kim; Buono and Eakin; and Noulas, Ray, and Miller examine both product specific economies of scale and economies of scope. These studies all define and calculate the measures of cost economies as discussed above.

Mester studies the California savings and loan industry using 1982 data. She defines the three outputs as mortgage loans, other loans, and cash, securities, other investments, etc. Mester estimates product specific elasticities of 1.19 for mortgage loans, -8.24 for other loans, and -

15.86 for cash, securities, other investments, etc. Only, the estimate on mortgage loans is statistically significantly different from one, suggesting that only mortgage loans exhibit product specific economies of scale. Mester estimates the economies of scope for the three outputs to be 1.06. This estimate was not significantly different from zero. Mester also found that capital and labor are substitutes. Mester concludes that no evidence of economies of scale is found in her study indicating natural monopolies do not exist in the California savings and loan industry.

Murray and White, and Kim used the same data set to examine economies of scale and of scope for British Columbia credit unions over the 1976 through 1977 period. Murray and White and Kim (MWK) use the intermediation approach as did Mester. MWK define three outputs; mortgage loans, other loans, and investments in excess of minimum liquidity requirements in the estimation of their translog cost function. Input categories defined were labor services, capital, demand deposits, and term deposits plus shares. The price variables were determined using much the same methods as Mester.

Kim estimates product specific economies of 1.01 for mortgage loans, .39 for other loans, and .98 for investments. Kim also estimates a global measure for economies of scope of .27 for the production of the three outputs simultaneously. Kim concludes that the finding of product-specific diseconomies of scale for nonmortgage loans in conjunction of the modest economies of scope raises doubt on the existence of natural monopolies in British Columbia credit unions. In addition, Murray and White find that labor and capital are substitutes and that homotheticity is not supported by the data.

Buono and Eakin use functional cost data and a hybrid translog cost function to examine economies of scale for 613 U.S. banks for 1985. The hybrid translog differs from the translog in that outputs are not logged while input prices are logged. This alleviates the problem caused by zero output. Buono and Eakin divide the data up to analyze economies of scope and scale in branch banking states and unit-banking states. Buono and Eakin use the intermediation approach to define three outputs: loans, investments, and transaction deposits. They define inputs to be capital, labor, time deposits, and purchased funds. The overall economies of scope is .642 for banks in branch banking states and .092 for banks in unit banking states. Buono and Eakin conclude that unit-state and branch-state banks have different cost structures. Economies of scope are statistically significant in branch-state banks but not in unit-state banks.

Noulas, Ray, and Miller (NRM) examine economies of scale using the intermediation approach in banks with total assets greater than $1 billion. NRM use 1986 Call Report data. NRM define 4 outputs as personal loans, real estate loans, commercial and industrial loans, and other investments. NRM define 4 inputs as interest bearing 1) deposits greater than $100,000 excluding CDs (deposits), 2) CDs greater than $100,000, federal funds, demand notes, and other borrowed money (funds), 3) labor, and 4) capital. NRM treat non-interest-bearing accounts as quasi-fixed assets.

NRM reject homothetic production of outputs. They also reject the notion of short-run and long-run constant returns to scale. They find that banks with assets between $1 and $3 billion exhibit

economies to scale, while diseconomies of scale arise in banks of $3 billion and larger. However, the economies and diseconomies of scale are not large. A problem with these findings is that no confidence interval is placed around these elasticity estimates. NRM find that all inputs are substitutes, with labor and capital having the largest degree of substitutability.

A common thread in all of these multiproduct cost studies is the use of the intermediation approach and the similar definitions of inputs. These studies reject the hypothesis of homothetic production technologies. Thus, the aggregation of output into a single commodity is inappropriate. Another common finding in these studies is that some evidence of economies of scale does exist for low levels of output, while diseconomies of scale exist for high levels of output. However, the statistical significance of these results is not all that strong. Each of these studies also find that global economies of scale are positive and exist, however, the estimates are not statistically significant.

Conclusions

The efficient size of financial institutions is an important issue for the remainder of the 1990s. With the recent waive of consolidation in the banking industry, an important issue is whether the mergers are further exhausting economies of scale, or whether merger activity is creating market power. This manuscript has examined the state of knowledge in bank efficiency from empirical parametric multiproduct cost concepts. Individuals interested in agricultural banking can learn a great deal from the banking literature as they begin to explore issues related to the efficiency of agricultural banks.

Many of the more recent banking studies have used the intermediation approach to defining bank activity rather than the production approach. The evidence in the finance literature is leaning towards the rejection of homothetic production functions. Thus, the use of a single output to represent bank output is inappropriate. In addition, a fruitful area of study will examine differences in cost surfaces in states with different banking laws. Finally, using longer time series of data will add robustness to many of the previous empirical studies. Most importantly, studies have not considered agricultural loans as a distinct output. Exploration of this part of the cost surface offers opportunities to expand the frontiers of knowledge for agricultural bankers.

References

Akridge, J.A. "Measuring Productive Efficiency in Multiple Product Agribusiness Firms: A Dual Approach." American Journal of Agricultural Economics 71(1989):116-25.

Akridge, J.A. and T.W. Hertel. "Measuring Productive Efficiency in Multiple Product Agribusiness Firms: A Dual Approach." American Journal of Agricultural Economics 68(1986):928-38.

Baumol, W.J., J.C. Panzar, R.D. Willig. Contestable Markets and Industry Structure New York, Harcourt Brace Jovanovich, Inc., 1982.

Benston, G.J., G.A. Hanweck, and D.B. Humphrey. "Scale Economies in Banking." Journal of Money, Credit, and Banking, 14(November 1982):435-56.

Berger, A.N., G.A. Hanweck, and D.B. Humphrey. "Competitive Viability in Banking." Journal of Monetary Economics, 20(1987):501-20.

Buono, M.J. and B.K. Eakin. "Branching Restrictions and Banking Costs." Journal of Banking and Finance 14(1990):1151-62.

Clark, J.A. "Economies of Scale and Scope at Depository Financial Institutions: A Review of Literature." Federal Reserve Bank of Kansas City Economic Review 73(September/October 1988):16-33.

Clark, J.A. "Estimation of Economies of Scale in Banking Using a Generalized Functional Form." Journal of Money, Credit and Banking 16(1984):53-68.

Cowing, T.G. and A.G. Holtmann. "Multiproduct Short-Run Hospital Cost Functions: Empirical Evidence and Policy Implications from Cross-Section Data." Southern Economic Journal 49(1983):637-53.

Deller, S.C., D.L. Chicoine and N. Walzer. "Economies of Size and Scope in Rural Low-Volume Roads." Review of Economics and Statistics, 70(1988):459-65.

Dowling, W.A. and G.C. Philippatos. "Economies of Scale in the U.S. Savings and Loan Industry: 1973-83." Applied Economics 22(1990):237-48.

Eakin, B.K., D.P. McMillen and M.J. Buono. "Constructing Confidence Intervals using the Bootstrap: An Application to a Multiproduct Cost Function." Review of Economics and Statistics 72(1990):339-44.

Efron, B. "Better Bootstrap Confidence Intervals." Journal of the American Statistical Association 82(1987):171-85.

Farrell, M.J. "The Measurement of Productive Efficiency." Journal of the Royal Statistical Society Series A 120(1957):253-81.

Gallant, A.R. "The Fourier Flexible Form." American Journal of Agricultural Economics 66(1984):204-8.

Gilligan, T.W. and M. Smirlock. "An Empirical Study of Joint Production and Scale Economies in Commercial Banking." Journal of Banking and Finance 8(1984):67-77.

Goldstein, S.J., J.E. McNulty, and J. Verbrugge. "Scale Economies in the Savings and Loan Industry Before Diversification." Journal of Economics and Business 39(1987):199-207.

Humphrey, D.B. "Why Do Estimates of Bank Scale Economies Differ?" Federal Reserve Bank of Richmond Economic Review 76(September/October 1990):38-50.

Kilbride, B.J., B. McDonald, and R.E. Miller. "A Reexamination of Economies of Scale in Banking Using a Generalized Functional Form." Journal of Money, Credit and Banking 18(1986):519-26.

Kim, H.Y. "Economies of Scale and Scope in Multiproduct Financial Institutions: Further Evidence from Credit Unions." Journal of Money, Credit and Banking 18(1986):220-6.

Mester, L.J. "A Multiproduct Cost Study of Savings and Loans." Journal of Finance 42(1987):423-45.

Murray, J.D. and R.W. White. "Economies of Scale and Economies of Scope in Multiproduct Financial Institutions: A Study of British Columbia Credit Unions." Journal of Finance 38(1983):887-901.

Nelson, R.W. "Branching, Scale Economies, and Banking Costs;" Journal of Banking and Finance 9(1985):177-91.

Noulas, A.G., S.C. Ray, and S.H. Miller. "Returns to Scale and Input Substitution for Large U.S. Banks." Journal of Money, Credit and Banking 22(1990):94-108.

Schroeder, T.C. "Economics of Scale and Scope for Agricultural Supply and Marketing Cooperatives." Review of Agricultural Economics 14(1992):93-104.

Shaffer, S. and E. David. "Economies of Superscale in Commercial Banking." Applied Economics 23(1991):283-93.

Shephard, R.W. Cost and Production Functions, Princeton, New Jersey: Princeton University Press, 1953.

White, H. "Using Least Squares to Approximate Unknown Regression Functions." International Economic Review 21(1980):149-70.