

Economies of Scale and Scope At Depository Financial Institutions: A Review of the Literature

By Jeffrey A. Clark

In recent years, changes in laws and regulations have greatly increased the opportunities for commercial banks and other depository financial institutions to expand their operations. Restrictions on interstate banking and intrastate branching have been liberalized in many states. In addition, limitations have been narrowed on the types of services depository institutions can offer.

While these changes have created new opportunities for individual depository institutions to grow, they have raised questions about the future structure of the banking industry. As some institutions expand and others fall prey to competitive pressures and decline or disappear, the industry's structure might come to be dominated by a small number of large diversified institutions. The market power of these institutions might allow them to keep loan rates too high and deposit rates too low, resulting in a misallocation of the

nation's financial resources. The potential for resource misallocation would likely be attenuated by competitive pressures from nondepository financial institutions and from nonfinancial firms. Nevertheless, the evolving structure of the banking industry remains a source of interest and potential concern for industry observers, regulatory agencies, and policymakers.

The industry's evolving structure will depend on what types of depository institutions can remain profitable over time. Among the primary determinants of profitability will be the extent that production economies and resultant cost reductions can be achieved as firms expand their operations. If extensive cost reductions are possible, large diversified firms will potentially be more profitable than small specialized institutions.

By studying production and cost conditions that have prevailed in the past, some insight can be gained into whether the increased opportunities for growth will allow cost reduction to be achieved. This article reviews the recent literature and concludes that, in general, large diversified depository institutions have not enjoyed a large cost advantage over smaller, more specialized institutions. The article's first section discusses

Jeffrey A. Clark is associate professor of finance at Florida State University and a visiting scholar at the Federal Reserve Bank of Kansas City. The views expressed in this article are those of the author and do not necessarily represent those of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

production economies and their role in influencing industry structure. The second section reviews the empirical literature on production economies at depository financial institutions. Several important issues and problems that arise in the estimation of production economies are examined in the third section. The last section summarizes the article and describes several policy implications that may be drawn from this literature.

Production economies

Two types of production economies may be achieved by individual firms in any industry—economies of scale, which are associated with firm size, and economies of scope, which relate to the joint production of two or more products.¹ Firms in an industry realize economies of scale if technology allows production costs to rise proportionately less than output when output increases. That is, economies of scale exist if per-unit or average production costs decline as output rises. Conversely, if average costs rise with output, diseconomies of scale are present. Economies of scope arise if two or more products can be jointly produced at a lower cost than is incurred in their independent production. Diseconomies of scope are present if joint production is more costly than independent production.

Industry structure is greatly influenced by the nature of production economies. If an industry's technology allows for both economies of scale and economies of scope, the industry will tend to be made up of large diversified firms.² These

¹ For an extensive discussion of economies of scale, see Scherer (1980).

² In the economics literature, these institutions would be termed competitively viable. More formally, a firm is defined as competitively viable if, in the long run, no other firm can produce a given product, or product mix, at a lower per-unit cost. To an economist the concept of cost means opportunity cost. Thus, the definition of competitive viability is inclusive of all revenue

firms will be able to produce at lower per-unit costs than smaller specialized firms and can potentially use this cost advantage to gain market share. Alternatively, if technology allows neither economies of scale nor scope, small specialized firms will tend to dominate the industry. A mixture of larger diversified firms and smaller specialized firms will develop in the absence of significant economies of scale and scope.

Economies of scale

There are two kinds of economies of scale. Economies that arise from increases in the production of individual products are called product-specific economies of scale. Economies associated with increases in all of a firm's outputs are referred to as overall economies of scale.

While the two types are synonymous for a single-product firm, both types of scale economies may be present for firms that produce more than one product. For multiproduct firms, overall economies of scale occur if total costs increase proportionately less than output when there is a simultaneous and equal percentage increase in each of the firm's products. With overall economies of scale, average costs decline as the firm expands production while maintaining a constant product mix.

Product-specific economies of scale are present if a decline in the per-unit cost of producing a specific product occurs as the output of that product increases. In principle, product-specific economies of scale for each product should be measured independently from the other products in the product mix. However, in practice such a measure is not meaningful since, under joint

and cost streams generated by alternative uses of the firm's assets. That is, if a firm's long-run costs are not at a minimum, there will be an incentive to increase profit by altering the level and/or mix of firm output.

production, it is generally impossible to change the output of one product while holding constant the output of the other products.³ In spite of this problem, an approximate measure of product-specific economies of scale has been proposed and used in the empirical literature. This measure is discussed in the box on page 27.

Economies of scope

There are two types of economies of scope, global and product-specific. To define global economies of scope, it is necessary to compare the costs of both joint production and separate production, assuming a given scale for each product. For a given product mix, if the total costs from joint production of all products in the product mix are less than the sum of the costs of producing each product independently, global economies of scope are present.

Product-specific economies of scope refer to economies that arise from the joint production of a particular product with other products. If production efficiency can be enhanced by adding a particular product to a given product mix, then product-specific economies of scope exist. That is, if the cost of producing a product independently from the other products in the product mix exceeds the cost of producing it jointly, product-specific economies of scope can be realized from joint production.

Product-specific economies of scope for a given product may result from joint production efficiencies with one or a large number of products in the mix. To determine which product pairs share jointness in production, cost complementarities between all pairs of products can be computed.

³ For expanded discussion of this problem, see Fuss and Waverman (1981).

A cost complementarity exists between two products if the marginal cost of producing one product declines when it is produced jointly with the other.

Sources of production economies at depository institutions

The literature on the theory of the firm has hypothesized numerous ways in which economies of scale and scope might arise in production. Making better use of specialized labor and capital and spreading fixed costs over large levels of output are usually cited as the predominant sources of economies of scale. Most economies of scope are thought to arise from the joint usage of a fixed resource.

Consistent with the theory of the firm, research on production by depository institutions often points to these important sources of both economies of scale and scope: specialized labor, computer and telecommunications technology, and information. For example, at small depository institutions, labor is unlikely to perform specialized functions. Tellers and loan officers probably process a variety of loan and deposit accounts since they are likely to be underutilized in handling specialized products. Their unspecialized labor is then a fixed input that can be shared in the production of a number of products, with the potential to create economies of scope. As these smaller institutions grow, they may be able to fully employ more specialized labor in producing some or all of their products. If the expertise of specialized tellers and loan officers results in the processing of a greater volume of deposit and loan accounts per unit of labor, then per-unit labor costs can be reduced through increased specialization. In this example, increased size may result in production efficiencies through the substitution of economies of scale for economies of scope.

The adoption of computer and telecommunica-

tions equipment can provide another basis for both economies of scale and scope at depository institutions. Despite the large set-up costs required, computer and other electronic funds transfer equipment can process a large volume of transactions at a small additional cost per transaction. As depository institutions increase the number of transactions of all types that can be performed by this equipment, it may be possible to reduce the per-unit cost of the firm as a whole as well as for individual products. Embracing this technology may provide a basis for both overall and product-specific economies of scale. In addition, any excess capacity of the equipment could be used to process other types of accounts at a small additional cost per transaction, thus realizing economies of scope.

Economies of scale and scope may also accompany information production. Before lending decisions can be made, credit information must be gathered and analyzed. Once gathered, however, this information can be reused in other lending decisions. Where the cost of reusing information is less than the independent cost of its production, reuse can help reduce the incremental costs of extending additional credit. If the information is reused to make similar loans to the same customer or to other customers in the same region or industry, it will provide a source of economies of scale. Alternatively, if the information can be used to make unrelated types of loans to the institution's customers, it may serve as a source of economies of scope.

A review of the empirical literature

Most of the evidence about the existence and extent of production economies at depository institutions comes from the empirical estimation of statistical cost functions. In developing these functions, researchers begin with the microeconomic principle that production costs depend on input prices and the level and composition of out-

put.⁴ After defining these variables, the researcher selects a statistical function to explicitly relate production costs to outputs and input prices. The most frequently selected statistical function is the transcendental logarithmic or translog function. This function is usually selected because it is flexible enough to yield both economies and diseconomies of scale at different output levels and to provide information on scope economies by incorporating interdependencies between products.⁵

Once the statistical function is selected and specified, the researcher estimates the parameters of the function using sample data. The estimated parameters and sample data are then used to construct empirical measures of the various types of scale and scope economies discussed in the previous section. A discussion of the most frequently used empirical measures is presented in the box on page 27. Technical statements of each measure are presented in Appendix B.

Empirical evidence

The 13 studies reviewed in this article attempted to estimate economies of scale and scope for credit unions, savings and loan associations, or commercial banks. Each study used a translog statistical cost function and employed similar measures of economies of scale and scope. The studies' results suggest four broad conclusions: First, overall economies of scale appear to exist

⁴ This functional relationship follows from the property of duality between the production and cost functions. When the statistical cost function is being estimated with cross-sectional data, it may be necessary to include other variables that may induce inter-firm variation in cost. Among the variables most commonly included are the number of branches and affiliation with a holding company.

⁵ An example of the general form taken by the translog function appears in Appendix B.

only at low levels of output with diseconomies of scale at large output levels. Second, there is no consistent evidence of global economies of scope. Third, there is some evidence of cost complementarities (product-specific economies of scope) in production. Finally, these results appear to be generally robust across the three types of institutions, as well as across different data sets and product and cost definitions.

Twelve of the 13 studies report significant overall economies of scale at relatively low levels of output (Table 1, column 2). Only Mester fails to find any evidence of scale economies, and then only for savings and loan associations below \$100 million of deposits.⁶ Only two studies, however, find significant overall economies of scale above \$100 million of deposits (Table 1, column 3). Moreover, the authors of one of these—Goldstein, McNulty, and Verbrugge—do not directly control for potential scope economies, and the authors of the other study—Benston, Hanweck, and Humphrey—report scale economies only for large branch banking organizations.⁷ Several authors report greater economies of scale among branch banking institutions, but when an augmented measure of overall economies of scale is employed to control for the interdependency between the number of offices and the number of accounts serviced, the cost advantage of branch banks seems to disappear.⁸

As already noted, it is not conceptually possible to measure product-specific economies of scale without ambiguities, so it may not be sur-

prising that only four of the 13 studies report evidence on this type of production economy.⁹ The results presented in these four studies do not support a conclusion of widespread product-specific economies of scale (Table 1, column 4). Both H.Y. Kim and Mester report product specific economies of scale for mortgage loans. However, H.Y. Kim and Gilligan, Smirlock and Marshall also report product specific diseconomies of scale for several products.¹⁰ Benston, Berger, Hanweck, and Humphrey report estimates of the marginal cost of production for five products by size class. However, they acknowledge that the negative marginal costs reported for some products are “implausible” and most likely indicate some type of estimation problem.¹¹

Eleven of the studies compute a measure of global economies of scope. However, only three report evidence of statistical significance for their measure. Further, in two of the three studies that report statistically significant global economies of scope, the statistical cost function that was estimated contained only two broadly defined products.¹² Only M. Kim reported statistically

⁶ See Mester (1987).

⁷ See Goldstein, McNulty, and Verbrugge (1987); and Benston, Hanweck, and Humphrey (1982).

⁸ See Appendix B for a presentation of the augmented measure of overall economies of scale used to control for the relationship between the number of offices and the number of accounts.

⁹ Appendix B presents several methods proposed by these authors for measuring product-specific economies of scale.

¹⁰ These products include nonmortgage loans, investment services, total loans and total deposits.

¹¹ The authors suggest that the most likely estimation problems are the presence of multicollinearity and the loss of degrees of freedom, both resulting from the large number of parameters that must be estimated when the translog function is used. See Benston et al. (1983).

¹² Gilligan, Smirlock, and Marshall (1984) include total deposit accounts and total loan accounts as the only two products in the cost function they estimate. Gilligan and Smirlock (1984) estimate two statistical cost functions, each with a pair of products. The product pairs employed in the two cost functions are, respectively, the total dollar amounts of demand and time deposits, and the total dollar amounts of total loans outstanding and total securities held.

TABLE 1
Summary of results of studies reviewed

Authors	Significant economies of scale			Significant economies of scope	
	Overall			Global	Cost complementarities
	Below \$100 million in deposits	Above \$100 million in deposits	Product specific		
Murray and White (1983)	yes	no	no measure	no	yes
H.Y. Kim (1987)	yes ^c	no	yes ^{o,j} no ^{p,j}	no ^k	no ^k
Mester (1987)	no	no	yes ^o	no	no
Goldstein, McNulty, & Verbrugge (1987)	yes ^a	yes ^a	no measure	no measure	no measure
LaCompte and Smith (1986)	yes ^b	no	no measure	no	yes(1978) no(1983)
Benston, Hanweck, & Humphrey (1982)	yes ^{d,f} no ^e	yes ^{d,f} no ^e	no measure	no measure ^l	no measure ^l
Benston, Berger, Hanweck, & Humphrey (1983)	yes	no	yes ^j	no	yes ^j
Gilligan and Smirlock (1984)	yes	no	no measure	yes ^m	no measure
Gilligan, Smirlock, & Marshall (1984)	yes	no	no ^{s,j}	yes ^m	no measure
M. Kim (1986)	yes	no	no measure	yes ⁿ no	yes ⁿ no
Lawrence and Shay (1986)	no ^g	no ^h	no measure	no ^o	yes ^o
Berger, Hanweck, & Humphrey (1987)	yes	no	no measure	no	yes
Kolari and Zardhooki (1987)	no ^e yes ^d	no ^e no ^{d,i}	no measure	no	yes

- Notes:
- a: Did not control for economies of scope.
 - b: Up to \$50 million in total deposits.
 - c: Reports diseconomies of scale to nonmortgage lending.
 - d: Denotes branch banking.
 - e: Denotes unit banking.
 - f: Reports diseconomies of scale if an augmented global scale economies measure is utilized.
 - g: Reports increasing returns to scale in 1980 and 1981 only.
 - h: No diseconomies of scale found in the upper two quartiles as high as \$2.5 billion in 1980 and 1981.
 - i: Up to \$100 million in total deposits.
 - j: Provides no statistical tests.
 - k: Reports scope economies but without tests of statistical significance.
 - l: Employed Divisia Index for output.
 - m: Test of nonjointness restrictions used only one pair of outputs.
 - n: Denotes a no-aggregation model.
 - o: Diseconomies of scope found between loans and investments.
 - p: For mortgage loans only.
 - q: Reports diseconomies for nonmortgage lending and investment services.
 - r: Reports computed marginal costs for selected products and arbitrarily chosen deposit size classification.
 - s: Reports diseconomies of scale for total loans and total deposits.

significant global economies of scope for a more disaggregated product mix.¹³ The last two columns of Table 1 summarize the estimates of global and product-specific (cost complementarities from joint production) economies of scope.

Although the empirical evidence does not support a conclusion of global economies of scope from joint production, many of the studies report some evidence of cost complementarities between pairs of products. When the translog function is estimated, evidence of a cost complementarity between any two products is given by a negative and statistically significant parameter estimate on the cross-product term between the two products. Table 2 lists all product pairs for which the estimated cross-product term is statistically significant. Inspection of this table indicates that some evidence of cost complementarities can be found in a number of studies and among a variety of different product pairs. The strongest evidence of cost complementarities occurs in the joint production of two product pairs: total loans and total deposits, and investments and mortgage loans.¹⁴ However, diseconomies of joint production were also reported between two related product pairs: investments and total loans, and total loans and total deposits for branch banks with total deposits below \$100 million.¹⁵

¹³ In his study of Israeli banks, Kim (1986) defined several alternative product mixes as combinations of four distinct products: demand deposits, foreign currency, loans, and securities. His results indicate that global economies of scope only occur when the four products appear separately in the cost function. Kim reports an absence of global economies of scope for all other combinations of these four products.

¹⁴ A cost complementarity between total loans and total deposits is reported in Berger, Hanweck, and Humphrey (1987); Gilligan, Smirlock, and Marshall (1984); and Lawrence and Shay (1986). A cost complementarity between investments and mortgage loans is reported in LaCompte and Smith (1986), and Mester (1987).

¹⁵ The diseconomy of the first type is reported in Lawrence and Shay (1986). The second type of diseconomy is reported in

Issues and problems

Several issues and problems may have influenced the results discussed in the preceding section. These issues and problems are both conceptual and methodological in nature. The problems tend to limit, but not eliminate, the usefulness of the empirical conclusions in drawing policy implications.

Defining bank costs and output

The banking literature is divided over the conceptual issue of the appropriate definition of bank output, and consequently on the related issue of defining bank costs. In general, researchers take one of two approaches.¹⁶ These alternative approaches are labeled the "intermediation approach" and the "production approach."¹⁷ No consensus has developed favoring one of the definitions over the other, and reasonable arguments have been made for both approaches.

Under the intermediation approach, depository financial institutions are viewed as producers of services related directly to their role as an intermediary in financial markets. That is, they are viewed as collecting deposits and purchasing funds to be subsequently intermediated into loans and other assets. In this approach, deposits are treated as inputs along with capital and labor. Those authors who adopt this approach generally define the institution's various dollar volumes of earning assets as measures of output. Also con-

Berger, Hanweck, and Humphrey (1987). Lawrence and Shay report an additional diseconomy of joint production between non-bank activities and total deposits.

¹⁶ The approaches taken in the 13 papers reviewed here appear in the second column of Appendix A.

¹⁷ Discussions of these two approaches can be found in a number of recent papers including Humphrey (1987); Mester (1987a); and Berger, Hanweck, and Humphrey (1987).

TABLE 2
Significant cost complementarities

<u>Output Pairs</u>	<u>Author(s)</u>	<u>Year(s)</u>	<u>Sign</u>
Consumer and mortgage loans	LaCompte and Smith	1978	negative
Investments and total loans	Gilligan and Smirlock	1973-78	negative
	Lawrence and Shay	1982	positive
Nonbank activity and total loans	Lawrence and Shay	1982	negative
Total deposits and total loans	Lawrence and Shay	1982	negative
	Gilligan, Smirlock, & Marshall	1978	negative
	Berger, Hanweck, & Humphrey	1983	negative*
Investments and mortgage loans	Mester	1982	negative
	LaCompte and Smith	1978	negative
Nonbank activity and investments	Lawrence and Shay	1982	negative
Total deposits and investments	Lawrence and Shay	1982	negative
Nonbank activity and total deposits	Lawrence and Shay	1982	positive
Time deposits and demand deposits	Gilligan and Smirlock	1973-78	negative

* Negative for branch banks with deposits > \$100 million in total deposits; positive for branch banks < \$100 million in total deposits.

sistent with this approach, costs are defined to include both interest expense and total costs of production.

The production approach, on the other hand, views depository institutions as producers of services associated with individual loan and deposit accounts. These account services are produced using capital and labor. Under this approach, it follows that the number of accounts of each type are the appropriate definitions of outputs. Total costs are defined exclusive of interest costs.

Conceptually, the intermediation and production approaches are very different. In reviewing the literature, it is surprising that the empirical results do not appear to be sensitive to the approach taken in defining outputs and costs. Why this should be the case is unclear. However, one possibility is that other issues, as discussed below, are more important.

Data

One of two types of data has been employed in nearly all recent attempts at estimating statistical cost functions for depository institutions. The data are drawn either from Call Report and financial statement data (as reported to the Federal Deposit Insurance Corporation, the Federal Savings and Loan Insurance Corporation, and the National Credit Union Share Insurance Fund), or from the Functional Cost Analysis (FCA) program conducted by the Federal Reserve System.

Each of these two sources of data offers advantages and disadvantages. An advantage of the FCA data is that they are constructed using simple cost accounting techniques to allocate costs among several distinguishable banking functions. In addition, these data include information on the number and average size of a variety of deposit and loan products. However, the generalization of the results obtained using FCA data to all depository institutions may be inappropriate for several reasons. Because the FCA program is

voluntary, subscribing banks might be either high-cost institutions interested in identifying areas for cost reduction or low-cost firms that place greater emphasis on cost control. Further, the FCA data are heavily skewed toward small banks.¹⁸ Finally, the procedures used to allocate costs are sometimes imprecise and may induce unknown bias in parameter estimates when the FCA data are used to estimate statistical cost functions.¹⁹

An advantage of Call Report and financial statement data is that they provide information on a much wider range of institutional size and impose uniform reporting requirements. The empirical results obtained using these data, therefore, should be more generally applicable. However, this source of data also imposes limitations. First, the absence of information on numbers of deposit and loan accounts and average account size make this source of data unsuitable for use under the production approach. Further, there is some evidence that the average account size and institution size are positively correlated. Thus, a failure to control for average account size under the intermediation approach may tend to overstate any finding of economies of scale. Second, data on some banking functions such as loan commitments, standby letters of credit, safety deposit and trust activity have only recently, if at all, been reported in these data. Finally, it is questionable whether financial statement data can be used to construct meaningful proxies for the input prices, given the high level of aggregation at which these data are reported.

¹⁸ As of 1986, only 490 banks participated in the program. Of this number, 416 were under \$200 million in total deposits.

¹⁹ In some cases, the allocations are made according to the judgment of the participating banker (e.g., wages and salaries). In other instances, the allocations are performed by computer algorithms developed for a representative bank using "experience factors" that are derived from previous data. For additional discussion of the allocation rules, see the *Introduction to Functional Cost Analysis: 1986 Average Banks*.

Level of aggregation and limitations of the translog functional form

Two closely related issues arise in the estimation of scale and scope economies: the appropriate level of aggregation and the suitability of the translog functional form for use with data from depository institutions. Theoretically, a measure of each distinct product offered by depository institutions should be included in the estimated function. However, the feasibility of doing this is usually limited by the availability of data and the use of a translog functional form. The larger the number of distinct products that are defined, the greater the likelihood that institutions included in the sample do not produce some of the products. Since the translog function expresses each input price and the output of each product in logarithmic form, the values of these variables must be strictly greater than zero. If a high level of disaggregation is chosen to increase the ability to identify jointness in production, then smaller and more specialized depository institutions will need to be deleted from the sample. Alternatively, if the level of aggregation in defining products is high enough to provide positive values for the output of all defined products for all institutions in the sample, then much of the information on efficiencies from joint production may be lost.²⁰

A second problem involving the level of disaggregation and the translog functional form arises in attempting to compute measures of product specific economies of scale and global economies of scope. The computation of these measures requires the assumption of a zero level of output for at least one of the products being produced. However, the translog cost function will always

yield zero total costs whenever the output of even one product is zero. To circumvent this problem, most researchers compute total costs by choosing an arbitrarily small but nonzero value for use in place of zero. This procedure has two drawbacks. First, the arbitrarily chosen value is usually well outside the bounds of the data. As a consequence, the confidence intervals around any computed values for these measures will be extremely wide. Second, the conventional measure of global economies of scope can be made to yield scope diseconomies. This result can be insured by replacing all zero outputs with a sufficiently small nonzero value.²¹

A third source of problems involving the level of disaggregation and the translog function arises from the number of parameters that must be estimated. As more products are defined and included in the statistical cost function, the number of parameters that must be estimated increases disproportionately.²² For depository institutions, the number of products that must be defined to yield any meaningful level of disaggregation is large. With the necessity of including linear, quadratic, and cross-product terms for all defined products and input prices, the likelihood of severe multicollinearity would appear to be high. In this case, it may not be possible to identify individual parameter estimates. Any statistical tests will be imprecise since the standard errors of the parameter estimates are likely to be large.²³

²¹ A thorough discussion of this problem can be found in Benston et al. (1983).

²² Mester (1987a) has noted that the addition of one input and one product to a translog function consisting of three inputs and three products increases the number of parameters that must be estimated from 28 to 45.

²³ The author of this article estimated a translog cost function with seven defined products using a sample of 190 commercial banks in the Denver SMSA in 1987. All of the included banks had nonzero values for each defined product. The products

²⁰ Kim (1986) reports evidence that suggests if product definitions are drawn too broadly, the resulting parameter estimates will be biased against the identification of significant economies of scope.

Other incentives for joint production

The concept of cost in economics is synonymous with opportunity cost, not accounting cost. Thus, in principle, the measurement of economies of scale and scope using a statistical cost function should attempt to measure the total costs of production in terms of opportunity costs rather than accounting costs. While technology may provide opportunities for the sharing of inputs, the decision to add product lines will depend ultimately on whether the additional product will increase after-tax, risk-adjusted returns. The focus on accounting costs results in the exclusion of any revenue and tax-related incentives for adding product lines—such as a reduction in earnings volatility from increased diversification—that are not rooted in production efficiencies and may even increase per-unit accounting costs.²⁴

appearing in the cost function included the dollar value of transactions deposits, time deposits, investments, real estate loans, installment loans, credit card loans, and commercial loans, respectively. Under the assumption that at the margin all banks in the market faced the same input prices, use of the translog cost function required the estimation of 36 parameters. The estimation of the translog cost function produced an adjusted R² of .9783 and an F-statistic of 245.782. However, of the 36 parameters only four were statistically significant at the 5 percent level of significance. The variance decomposition collinearity diagnostics provided in SAS produced “high” condition indices for all but five of the variables. Further, there were numerous instances in which variables with high condition indices contributed strongly (exhibited a variable proportion greater than 0.5) to the variances of two or more variables. These results are indicative of a severe multicollinearity problem.

²⁴ Other incentives may include greater use of off-balance sheet activities to avoid regulatory taxes imposed by risk-based capital requirements and deposit insurance premiums, and joint customer demand for banking services that arise from a desire to reduce transactions costs. See Baer and Pavel (1988) for a recent analysis of the regulatory tax imposed by minimum capital requirements and deposit insurance premiums.

Summary

Care should be exercised in attempting to use the existing empirical literature as a sole basis for policy. At present, no systematic attempts have been made at conducting a sensitivity analysis of the empirical results to the issues and problems discussed above. Further, it is difficult to assess the severity of these problems by examining the existing literature because differences among studies are sufficiently large to prevent drawing conclusions on specific issues.

Finally, the studies reviewed in this article predate the granting of new securities, insurance, mutual funds, and other powers to depository institutions and therefore cannot be used to draw inferences about their likely impact on costs. This is particularly true since the size of any impact will depend importantly upon whether the new powers are granted directly to institutions or can be offered only through affiliates of bank holding companies.

Conclusions and policy implications

A review of the empirical evidence presented in 13 separate studies of economies of joint production for depository institutions yields several tentative results. First, the empirical evidence appears to support a conclusion of significant overall economies of scale only for depository institutions of relatively small size—less than \$100 million in total deposits. Second, the empirical evidence does not appear to support a conclusion of global economies of scope. Third, there appears to be some evidence of economies in joint production among specific pairs of products that might be offered by depository institutions.

The three results listed above suggest several tentative policy conclusions. Taken together, the evidence implies that the smallest, most specialized of depository institutions may be at a cost disadvantage relative to larger, more diversified

institutions. These smaller institutions are likely to be faced with the necessity of increasing both the scale and scope of their operations to remain cost competitive. Failure to achieve sufficient growth and to exploit available cost complementarities may drive these depository institutions from the market or cause them to be absorbed by other more cost-efficient institutions. However, the evidence also suggests that once overall scale economies have been exhausted, there will still be opportunities for the smaller, less diversified depository institutions. The absence of strong global economies of scope, combined with evidence of several cost complementarities, will probably provide a number of market niches for these smaller institutions.

From a policy perspective, the absence of a cost advantage for the largest, most diversified depository institutions appears to minimize any concern that the banking industry will be dominated by a few large depository financial institutions. The

lifting of restrictions on interstate banking and intrastate branching might help consolidate resources in states that have prohibited or severely limited branch banking by permitting small banks to achieve a more efficient scale of production. The absence of significant scope economies suggests, however, that the lifting of these restrictions is unlikely to require significant adjustment in product mix.

In light of the issues and problems raised in this article, there is ample room for more research. Future efforts should address questions like these: Is there a better statistical function for use in measuring economies of scale and scope than the translog cost function? What is the appropriate level for the disaggregation of output for depository institutions? What is the best way to broaden the focus to include incentives for joint production? And, as new powers are granted to depository institutions, how will this affect their production efficiencies?

Empirical Measures of Production Economies

Researchers have developed empirical measures for both economies of scale and economies of scope. Overall economies of scale are typically measured by computing the sum of the output cost elasticities of individual products. The output cost elasticity for a product is the percentage change in production costs that occurs for a given percentage change in the output of the product. And, the sum of the individual output cost elasticities is equivalent to the percentage change in costs that results from an equal percentage change in the output of all products. When this measure of overall economies of scale is equal to one at a given level of overall output, there are constant returns to scale. Thus, no additional production efficiencies can be achieved in this range of production. If this measure of overall scale economies

is significantly less than one, then there are increasing returns to scale and production efficiencies will be realized in this range of production. Conversely, if this measure is significantly greater than one, there are decreasing returns to scale and production inefficiencies will be realized.

While product-specific economies of scale cannot be measured without ambiguities, an approximate measure has been proposed and utilized in several cost studies. This measure makes use of the theoretical relationship between the marginal cost, average cost, and economies of scale. Where the marginal cost of producing a product is less than average cost at a given level of output, average cost is declining in that range of output, implying economies of scale. Conversely, when

marginal cost is greater than average cost, average cost is increasing, implying diseconomies of scale. To approximate this relationship in a multi-product setting, a new cost concept labeled "average incremental cost" (AIC) is utilized. AIC is defined as the addition to total cost of producing a specific level of a product as opposed to not producing it at all, divided by the level of output of the product. Then the AIC can be expressed as a ratio to the marginal cost of producing this level of output. If this ratio is greater than one, this is viewed as evidence of product-specific economies of scale for the range of output levels between zero and the level at which AIC and MC are evaluated, since it implies that average costs are declining. If the ratio is less than one, product-specific diseconomies of scale is implied.

Global economies of scope are measured by computing the cost differential that would arise between the independent and joint production of specific output levels of all products. This cost differential is then generally scaled by dividing by the total costs of joint production. This measure will have a value greater than zero when there are global economies of scope, and a negative value when diseconomies are present.

As an alternative to computing the preceding measure, researchers have demonstrated that a sufficient condition for global economies of scope is the existence of cost complementarities among all pairs of products in the product mix. A cost complementarity occurs when the marginal cost of producing one product declines with an increase in the level of production of another.

Product-specific economies of scope are measured in several alternative ways. One common

measure is to compute the cost increase or decrease that arises from producing a specific product both independently from, and jointly with, the remaining product mix and expressing it as a percentage of the costs of joint production. If this ratio is greater than one, product-specific economies of scope are implied. If the ratio is less than one, diseconomies of scope exist.

Other alternative ways of identifying a cost complementarity between any two products in the product mix involve an assessment of how joint production of two products affects the marginal cost of producing each product. When parameter estimates from a translog statistical cost function are used, it can be shown that a necessary condition for the marginal cost of producing a product to decline with an increase in the production of a second product, referred to here as a pairwise cost complementarity, requires their cross-product term to be negative and statistically different from zero. However, while a negative cross-product term is consistent with the existence of a cost complementarity, it is not sufficient. Any reduction in marginal costs from the joint production of the two products may be offset by rapidly rising marginal costs from one or both of the two products. When the translog function is estimated, it can be shown that a sufficient condition for a cost complementarity between two products requires that the cross-product term not only be negative but also greater in absolute value than the product of the output elasticities of the two products being considered. A statistical test of this condition (test of nonjointness) is carried out by testing the parameter restrictions that would be required for nonjointness in the production of the two products.

Appendix A

Summary of Studies Reviewed

<u>Authors</u>	<u>Approach</u>	<u>Data</u>	<u>Outputs</u>	<u>ECSCA*</u>	<u>ECSCO*</u>
Murray and White (1983)	Intermediation	61 Canadian Credit Unions (1976-77)	y_1, y_2, y_{14}	OSA(1)	PSSO(2)
H. Y. Kim (1986)	Intermediation	61 Canadian Credit Unions (1976-77)	y_1, y_2, y_{14}	OSA(1) PSSA(1)	GSO(1) GSO(2) PSSO(4)
Mester (1987)	Intermediation	149 Calif. S&Ls (1982)	y_1, y_3, y_{14}	OSA(1)	GSO(1) PSSO(1)
Goldstein, McNulty, & Verbrugge (1987)	Production	FSLIC Insured S&Ls (1978-81)	y_{13}	OSA(1)** OSA(2)**	No measure
LaCompte & Smith (1986)	Intermediation	S&Ls Ninth Dist. FHLBB (1978-83)	y_1, y_2, y_3	OSA(1)	PSSO(2)
Benston, Hanweck & Humphrey (1982)	Production and Intermediation	FCA Data (1975-78)	Divisia Index	OSA(1)**	No measure
Benston, Berger, Hanweck & Humphrey (1983)	Production	FCA Data [deposits less than one billion] (1978)	y_4, y_5, y_6, y_7, y_8	OSA(1) OSA(2)	PSSO(3) PSSO(4)
Gilligan & Smirlock (1984)	Production	Financial Statement Data, 2700 banks (1973-78)	y_3, y_4, y_5, y_9	OSA(1)	PSSO(3)
Gilligan, Smirlock, & Marshall (1984)	Production	FCA Data (1978)	y_9, y_{10}	PSSA(2)	PSSO(3)
M. Kim (1986)	Intermediation	17 Israeli Banks (1979-82)	y_3, y_4, y_9, y_{11}	OSA(1) OSA(2)	GSO(1) PSSO(1)
Lawrence & Shay (1986)	Intermediation	FCA Data (1979-82)	y_3, y_4, y_9, y_{12}	OSA(1)	PSSO(4)
Berger, Hanweck & Humphrey (1987)	Production	FCA Data (1983)	y_4, y_5, y_6, y_7, y_8	OSA(1) EPSA	GSO(1) EPSUB
Kolari & Zardhooki (1987)	Production	FCA Data (1979-1983)	$y_3, y_4, y_5, y_9, y_{10}$	OSA(1)	GSO(1) PSSO(2)

Notes: *See Appendices A and B for definitions of the abbreviations for the measures of economies of scale and scope employed in this table.

** indicates the use of a Divisia Index for output. Other outputs are denoted as follows: y_1 = mortgage loans; y_2 = consumer loans; y_3 = investments; y_4 = demand deposits; y_5 = time deposits; y_6 = real estate loans; y_7 = commercial loans; y_8 = installment loans; y_9 = total loans; y_{10} = total deposits; y_{11} = foreign currency; y_{12} = nonbank activities; y_{13} = total assets; and y_{14} = other loans.

Appendix B Empirical Measures of Economies of Scale and Scope

I. TRANSLOG STATISTICAL COST FUNCTION

$$\ln TC = B_0 + \sum_i B_i \ln y_i + \sum_k C_k \ln p_k + (1/2) \sum_i \sum_j D_{ij} \ln y_i \ln y_j \\ + (1/2) \sum_k \sum_l E_{kl} \ln p_k \ln p_l + \sum_i \sum_k F_{ik} \ln y_i \ln p_k + e,$$

where \ln denotes the logarithm; $y_i (i=1, \dots, m)$ denotes the i th output; $p_k (k=1, \dots, n)$ denotes the k th input price; $B_0, B_i, C_k, D_{ij}, E_{kl}, F_{ik}$ are the parameters to be estimated and e represents the random error term.

II. OVERALL ECONOMIES OF SCALE

A. Overall or Plant Economies of Scale

$$OSA(1) = \sum_i \frac{\partial \ln TC}{\partial \ln y_i} = \sum_i \epsilon_i, \text{ where } \epsilon_i$$

is the output cost elasticity for product i . $OSA(1) < 1$ indicates overall economies of scale. $OSA(1) > 1$ indicates overall diseconomies of scale.

B. Augmented or Firm Economies of Scale

$$OSA(2) = \sum_i \frac{\partial \ln TC}{\partial \ln y_i} + \frac{\partial \ln TC}{\partial \ln B} \times \frac{\partial \ln B}{\partial \ln y_i},$$

where B is the number of branches operated by the depository institution. $OSA(2) < 1$ indicates overall economies of scale. $OSA(2) > 1$ indicates overall diseconomies of scale.

III. PRODUCT-SPECIFIC ECONOMIES OF SCALE

A. Average Incremental Costs

$$PSSA(1) = [(IC_i/TC)/\epsilon_i], \text{ where } \epsilon_i = \frac{\partial \ln TC}{\partial \ln y_i}$$

$TC = C(y_1, \dots, y_m)$ and $IC_i = [C(y_1, \dots, y_m) - C(y_1, \dots, y_{i-1}, 0, y_{i+1}, \dots, y_m)]$. $PSSA(1) > 0$ indicates product-specific economies of scale for product y_i . $PSSA(1) < 0$ indicates product-specific diseconomies of scale for product y_i .

B. Declining Marginal Cost

$$PSSA(2) = \frac{\partial^2 TC}{\partial y_i^2} = \left(\frac{TC}{y_i^2}\right) \left[\frac{\partial \ln TC}{\partial \ln y_i} + \left(\frac{\partial \ln TC}{\partial \ln y_i}\right) \left(\frac{\partial \ln TC}{\partial \ln y_i} - 1\right) \right].$$

If $PSSA(2) < 0$ then marginal costs of product y_i are declining. This implies product-specific economies of scale for product y_i . $PSSA(2) > 0$ implies increasing marginal costs and product-specific diseconomies of scale for product y_i .

IV. EXPANSION PATH SCALE ECONOMIES

$$EPSA = \left\{ \sum_i \left[\left(\frac{y_i^B - y_i^A}{y_i^B} \right) \left(\frac{C(y_i^B)}{C(y_i^A)} - 1 \right) \right] \right\} \times \frac{\partial \ln TC^B}{\partial \ln y_i},$$

where y_i denotes the level of output of product i produced by small Firm A or large Firm B. $C(\)$ denotes the total cost of producing level y_i of product i by each type of firm. If $EPSA < 1$ this implies economies of scale along an expansion path including firms A and B. If $EPSA > 1$ this implies diseconomies of scale along this expansion path.

V. GLOBAL ECONOMIES OF SCOPE

A. Global Economies of Scope

$$GSO(1) = \{ [C(y_1, o, \dots, o) + \dots + C(o, \dots, o, y_m)] - C(y_1, \dots, y_m) \} / C(y_1, \dots, y_m),$$

where $C(\)$ denotes the total costs of production. If $GSO(1) > 0$ then there are global economies of scope. If $GSO(1) < 0$ there are global diseconomies of scope.

B. Disjoint-Group Economies of Scope

$$GSO(2) = \{ [C(y_1, \dots, y_j) + C(y_{j+1}, \dots, y_m)] - C(y_1, \dots, y_m) \} / C(y_1, \dots, y_m), \text{ where}$$

$C(\)$ denotes the total costs of production. $GSO(2) > 0$ denotes economies of scope in production. $GSO(2) < 0$ denotes diseconomies of scope.

VI. PRODUCT-SPECIFIC ECONOMIES OF SCOPE

A. Product-Specific Economies of Scope

$$PSSO(1) = \{ [C(y_1, \dots, y_{i-1}, o, y_{i+1}, \dots, y_m) + C(o, \dots, o, y_1, o, \dots, o)] - C(y_1, \dots, y_m) \} / C(y_1, \dots, y_m)$$

where $C(\)$ denotes the total costs of production. $PSSO(1) > 0$ implies product-specific economies of scope. $PSSO(1) < 0$ implies product-specific diseconomies of scope.

B. Cost Complementarities

$$PSSO(2) = \frac{\partial^2 TC}{\partial y_i \partial y_k} = \left(\frac{TC}{y_i y_k} \right) \left[\frac{\partial^2 \ln TC}{\partial \ln y_i \partial \ln y_k} + \left(\frac{\partial \ln TC}{\partial \ln y_i} \right) \left(\frac{\partial \ln TC}{\partial \ln y_k} \right) \right]$$

$PSSO(2) < 0$ implies that an increase in the level of production of product y_k reduces the marginal cost of producing product y_i . Thus $PSSO(2) < 0$ implies product-specific economies of scope between products y_i and y_k . Conversely, $PSSO(2) > 0$ implies product-specific diseconomies of scope between products y_i and y_k .

C. Test of Nonjointness

From $PSSO(2)$, nonjointness implies $(\partial^2 TC / \partial y_i \partial y_k) = 0$. At any nonzero level of production of y_i and y_k , $(TC / y_i y_k) > 0$. Therefore, nonjointness requires

$$PSSA(3) = \left[\frac{\partial \ln TC}{\partial \ln y_i \partial \ln y_k} + \left(\frac{\partial \ln TC}{\partial \ln y_i} \right) \left(\frac{\partial \ln TC}{\partial \ln y_k} \right) \right] \leq 0.$$

From the translog this implies the restrictions that

$$[D_{ij} + \epsilon_i \times \epsilon_k] = 0, \text{ where}$$

$$\epsilon_i = \frac{\partial \ln TC}{\partial \ln y_i} = B_i + \sum_j D_{ij} \ln y_j + \sum_k F_{ik} \ln p_k.$$

The parameter restrictions can be imposed and a likelihood ratio test of the restrictions can be conducted.

D. Pairwise Cost Complementarities

A necessary condition for $(\partial^2 TC / \partial y_i \partial y_k) < 0$, is that the value of $(\partial \ln TC / \partial \ln y_i \partial \ln y_k) < 0$. This follows because, as in $PSSO(3)$, $(TC / y_i y_k) > 0$. Further, from theory, $MC_i = (\partial TC / \partial y_i) > 0$, so that $(\partial \ln TC / \partial \ln y_i) = (\partial TC / \partial y_i) (y_i / TC) > 0$. Therefore, a necessary condition for the existence of a cost complementarity between products y_i and y_k , when estimating the translog cost function, is

$$PSSO(4) = \frac{\partial^2 \ln TC}{\partial \ln y_i \partial \ln y_k} = D_{ik} < 0.$$

VII. EXPANSION PATH SUBADDITIVITY

$$EPSUB = \{ [C(Y^A) + C(Y^D) - C(Y^B)] / C(Y^B) \},$$

where $Y^A = (y_1^A, y_2^A, \dots, y_m^A)$ is the product-mix of small firm A, $Y^B = (y_1^B, y_2^B, \dots, y_m^B)$ is the product-mix of large firm B, and $Y^D = (Y^B - Y^A)$; $y_i^B \geq y_i^A \geq 0 \forall i$. $EPSUB > 0$ implies a cost advantage for large firm A.

References

- Baer, Herbert L. and Christie A. Pavel. "Does Deregulation Drive Innovation?" *Economic Perspectives*. Federal Reserve Bank of Chicago, March/April 1988, 3-16.
- Bailey, Elizabeth E. and Ann F. Friedlaender. "Market Structure and Multiproduct Industries." *Journal of Economic Literature*. September 20, 1982, 1024-1048.
- Baumol, William J., John C. Panzar, and Robert D. Willig. *Contestable Markets and Industry Structure*. New York, Harcourt Brace Jovanovich, Inc., 1982.
- Benston, George J., Gerald A. Hanweck, and David Humphrey. "Scale Economies in Banking." *Journal of Money, Credit, and Banking*, 14, Part 1, November 1982, 435-456.
- Benston, George J., Allen N. Berger, Gerald A. Hanweck, and David Humphrey. "Economies of Scale and Scope." *Proceedings of a Conference on Bank Structure and Competition*. Chicago, Federal Reserve Bank of Chicago, 1983.
- Berger, Allen N., Gerald A. Hanweck, and David B. Humphrey. "Competitive Viability in Banking." *Journal of Monetary Economics*, 20, 1987, 501-520.
- Functional Cost Analysis: 1986 Average Banks*. Federal Reserve Banks, 1986.
- Gilligan, Thomas W. and Michael L. Smirlock. "An Empirical Study of Joint Production and Scale Economies in Commercial Banking." *Journal of Banking and Finance*, 8, 1984, 67-77.
- Gilligan, Thomas W., Michael Smirlock, and William Marshall. "Scale and Scope Economies in the Multi-Product Banking Firm." *Journal of Monetary Economics*, 13, 1984, 393-405.
- Goldstein, Steven J., James E. McNulty, and James Verbrugge. "Scale Economies in the Savings and Loan Industry Before Diversification." *Journal of Economics and Business*, 1987, 199-207.
- Humphrey, David B. "Cost Dispersion and the Measurement of Economies in Banking." *Economic Review*, Federal Reserve Bank of Richmond, May/June 1987, 24-38.
- Kim, H. Youn. "Economies of Scale and Scope in Multiproduct Financial Institutions: Further Evidence from Credit Unions." *Journal of Money, Credit, and Banking*, 18, May 1986, 220-226.
- Kim, Moshe. "Banking Technology and the Existence of a Consistent Output Aggregate." *Journal of Monetary Economics*, 18, 1986, 181-195.
- Kolari, James and Asghar Zardhooki. *Bank Cost, Structure, and Performance*. Lexington, Mass., D.C. Heath, 1987.
- LaCompte, Richard L.B. and Stephen D. Smith. "The Impact of Regulation on Cost Structures: The Case of the Savings and Loan Industry." Unpublished manuscript, September 1986.
- Lawrence, Colin and Robert Shay. "Technology and Financial Intermediation in a Multiproduct Banking Firm: An Econometric Study of U.S. Banks, 1979-82." C. Lawrence and R. Shay, eds., *Technological Innovation, Regulation, and the Monetary Economy*. Cambridge, Mass., Ballinger, 1986.
- Mester, Loretta J. "Efficient Production of Financial Services: Scale and Scope." *Business Review*, Federal Reserve Bank of Philadelphia, January/February 1987a, 15-25.
- Mester, Loretta J. "A Multiproduct Cost Study of Savings and Loans." *Journal of Finance*, 42, June 1987, 423-445.
- Murray, John D. and Robert W. White. "Economies of Scale and Economies of Scope in Multiproduct Financial Institutions: A Study of British Columbia Credit Unions." *Journal of Finance*, 38, June 1983, 887-901.
- Scherer, F.M. *Industrial Market Performance and Economic Performance*. Chicago, Rand McNally, 2nd edition, 1980.