FRBSF ECONOMIC LETTER

1997-06 | February 28, 1997

Efficiency of U.S. Banking Firms – An Overview

Simon Kwan

Introduction

Bank managers, policymakers, and bank investors all are concerned with how efficiently a bank uses its labor and capital inputs to produce the cluster of financial products. Is a bank using the right level and mix of inputs? Is it producing the right output mix? Is it operating at the optimal scale? Since a bank's profitability is directly driven by its operating efficiency, the bank manager who is hired by the shareholders to maximize their wealth should be concerned with these questions; an under-performing bank would be penalized by the capital market by depressing its share price and subjecting it to takeover. Banking supervisors also pay attention to operating efficiency because a bank's safety and soundness is dependent on the quality of its management. In fact, managerial quality is one of the five components of the so-called CAMEL (Capital, Assets, Management, Earnings, and Leverage) rating in the bank examination process. In an attempt to explore the economics underlying these questions, economists have applied advance production theory and sophisticated econometric techniques to estimate bank efficiency, and a great deal of progress has been made in the past twenty years. This *Economic Letter* provides an overview of the findings of the research on the efficiency of U.S. banks.

Scale Efficiency

One dimension of banking efficiency that draws a lot of researchers' attention is scale efficiency, which refers to the relationship between a firm's per unit average production cost and production volume. For example, a firm that increases its production of outputs can see its unit cost of production decline, because the cost of some elements of inputs are fixed, such as administrative and overhead expenses. This is called economies of scale. However, diseconomies of scale also are possible if the average production cost starts to rise with outputs beyond a certain volume of production. Scale diseconomies

may arise because it may be more costly to manage a very large firm, or because the management of a very large firm becomes entrenched and therefore concerned more about maximizing their own welfare than that of shareholders. If the average cost curve is U-shaped–due to economies of scale at a low output level and diseconomies of scale at a high output level–it implies that there is an optimal scale of production at which the per unit average production cost is minimized.

For banking, prior research on scale efficiency found that the average cost curve was a relatively flat U-shape (see the survey by Humphrey 1990). However, the location of the optimal production scale remains unclear. In the U.S., there are a small number of very large banks that operate alongside a large number of smaller banks. The huge disparity in sizes among U.S. banking firms makes it difficult to estimate the average cost curve precisely. Furthermore, in general, the product mix at large banks is very different from the product mix at smaller banks, suggesting that it may not be appropriate to combine large banks with smaller ones when doing cost analysis. Consistent with this, studies that exclude large banks reported that average costs appear to be minimized at about $100 million to $300 million in total assets, while research that focused exclusively on large banks found that their optimal scale seems to be between $2 and $10 billion in total assets.

Recently, a major wave of consolidation has swept through the U.S. banking industry, affecting firms in almost all size classes. With respect to mergers between large institutions, the marriage of Chase Manhattan Corporation and Chemical Banking Corporation created the largest banking company ($297 billion in total assets) in the U.S., while the acquisition of First Interstate Bancorp by Wells Fargo & Company ($108 billion in combined total assets) represented the largest bank merger (in a deal valued at $11.6 billion). In both cases, cost reduction is cited as a very important motivation for the merger. However, if the average cost curve is truly U-shaped, the fact that the already very large banks were getting much larger would be inconsistent with cost minimization.

Apparently, bankers think that the optimal scale is much bigger than prior research suggests. This can be justified by the recent changes in the banking industry. First, the passage of the Riegle-Neal Interstate Banking and Branches Efficiency Act, which permits interstate branching effective 1997, may have changed the underlying production function of banking firms. Since banking firms do not need to set up separately capitalized banks in different states to deliver their financial products, it would be less costly to expand outputs. Second, the rapid change in information technology and the sharp drop in the cost of computers may have shifted the average cost curve. Third, banking firms have successfully cultivated alternative delivery channels for certain financial products, for example, through supermarkets, telephone lines, and the Internet, which may permanently alter the cost of providing certain banking products. As a result of fewer geographic barriers, rapid technological breakthroughs, and advances in distribution systems, it is quite plausible that the optimal scale of production may have increased. Moreover, the largest U.S. banking firm today is still not even in the worldwide top ten largest banks list based on 1995 total assets, suggesting that there may be room for further consolidation among large U.S. banks. Nevertheless, one cannot rule out the alternative hypothesis that entrenched managers use mergers and acquisitions to increase the size of the firm with the objective of enhancing their own welfare.

X-efficiency

While scale efficiency is about whether a banking firm has the right size, a potentially more important question is whether a firm produces as efficiently as it possibly can, given its size. This question can be answered by measuring X-efficiency, which in technical terms refers to deviations from the cost-efficient frontier that depicts the lowest production cost for a given level of output. X-efficiency stems from technical efficiency, which measures the degree of friction and waste in the production process, and allocative efficiency, which measures whether the right levels of various inputs are used.

A number of studies have investigated the X-efficiency of U.S. banks (see the survey by Berger, Hunter,

and Timme 1993). Most of the research done to date focused on the following question: Given the level of outputs and the price of capital, labor, and funds faced by the bank, does the bank minimize its total operating costs? If the observed total operating cost is higher than the minimized cost, the difference represents the bank's X-efficiency.

Researchers find that X-efficiency appears to vary substantially across banks. On average, the deviation from the minimum cost is found to be quite large, in the neighborhood of 20 to 25 percent of total costs, and it seems to dominate the effect of scale inefficiency. The findings suggest that for an average bank, the biggest room for efficiency gains lies in improving its operating efficiency, that is, doing things right, rather than on scale efficiency, that is, being the right size.

After identifying X-efficiency in banking firms, it is important to find out what may cause this type of inefficiency. While we do not have a conclusive answer to this question, preliminary evidence suggests that X-efficiency seems to be related to at least three things. First is size; on average, operating costs of larger banks are found to be closer to their minimum cost curve than those of smaller banks to their respective cost frontier. To the extent that large banks, which operate in metropolitan markets, face stronger competition than smaller banks, many of which operate in suburban or rural areas, it is not surprising that competition forces large banks to operate more efficiently. Second is risk-taking; inefficient firms are found to take on a higher level of risk. It is possible that managers of inefficient banks make loans with higher yields and higher risks in an attempt to compensate for their operating inefficiency. Furthermore, an inefficient firm, which has lower market valuation, has less to lose in taking a risky gamble than an efficient firm. Third is financial condition; banking firms with relatively more problem loans (that is, past due loans and non-accrual loans), tend to be less efficient than those with fewer problem loans. The correlation between poor asset quality and inefficiency may be an indication of poor management, or a direct consequence of the tendency of inefficient firms to make risky loans; alternatively, it may reflect the high operating cost of managing problem loans.

However, several caveats are in order. First, there are a number of methods of estimating X-efficiency, and the results are quite sensitive to the choice of method. Not only is the level of X-efficiency found to vary from method to method, but the ranking of X-efficiency of individual banks also is found to be somewhat inconsistent across methods. Second, by focusing solely on the cost function, researchers implicitly assumed that banking products are homogeneous. Since the cost function approach fails to trade off between product quality and production cost, a banking product that has more value-added to the customer but is more costly to produce will be treated as cost inefficient under this approach, despite the fact that a customer may be willing to pay more for a high quality product. Whereas the cost function approach is acceptable for analyzing commodity-type banking products, it is unacceptable when product differentiation is significant. To allow for variations in product quality, researchers have started to move away from the cost function approach and towards the revenue or profit function approach. Third, X-efficiency analyses were done in a static setting. They failed to capture changes in the regulatory environment and in the marketplace, which may have altered the underlying production process and the associated cost function. Based on these considerations, the research finding in X-efficiency studies should be interpreted with caution. These results are best treated as indicative rather than definitive, assisting decisionmakers in identifying potential problems instead of rendering judgment.

Conclusion

In the U.S. banking industry, researchers detected economies of scale, though the optimal scale appears to differ for large banks and smaller banks due to their very different product mix. They also suggest that the optimal size of large banks is well below that of today's megabanks. However, that finding may be obsolete given the dynamics of the banking industry. With significant regulatory changes, rapid technological breakthroughs, and constant product innovations, the right size for yesterday's environment may no longer be optimal today, much less for the future. More importantly, researchers

find that X-efficiencies, that is, deviations from the cost-efficient frontier, on average are large and dominate scale efficiency. Hence, it may be more fruitful to achieve efficiency gains by doing things right rather than by searching for the right scale.

Simon Kwan
Economist

References

Berger , A. N., W. C. Hunter, and S. G. Timme. 1993. “The Efficiency of Financial Institutions: A Review and Preview of Research Past, Present, and Future.” *Journal of Banking and Finance* 17, pp. 221-249.

Humphrey, D.B. 1990. “Why Do Estimates of Bank Scale Economies Differ?” Federal Reserve Bank of Richmond *Economic Review* 76 (Sept./Oct.), pp. 38-50.

Subscribe RSS Feed Share

More Economic Letters