



Scale Economies in Banking and Financial Regulatory Reform¹

Loretta J. Mester

Federal Reserve Bank of Philadelphia
and The Wharton School, University of
Pennsylvania

Global financial markets will be shaped for years to come by the regulatory reforms being implemented in response to the recent financial crisis. In my view, two key principles should guide reform efforts. First, reforms should take into account the incentives they create and their longer-run consequences. Second, reforms should harness market forces, not work against them.

U.S. policymakers have sought to foster stability by lowering the probability of a crisis and by reducing costs imposed on the rest of the economy when a shock hits the financial system. An important part of their deliberations has concerned financial firms deemed too big to fail or too interconnected to fail. I believe that, ironically, the United States will have a more stable financial system if failing firms are permitted to fail instead of being rescued.

Policymakers therefore need a way to allow a financial firm—of any size—to fail without precipitating a crisis. For this, a realistic “resolution” mechanism—a means of restructuring or dissolving a firm’s assets and liabilities—must be created. A credible mechanism must impose losses on creditors as well as shareholders and do it in a consistent manner so that stakeholders expect this imposition and have incentive to take adequate precautions against failure. The mechanism should be transparent and rule-based, giving regulators less discretion, not more.²

A related issue is how to deal with large or interconnected financial firms *before* they get into financial trouble. There has been a striking amount of consolidation in the banking industry in the United States and abroad over the past 30 years, and it has led to some very large banks. In the United States, the number of commercial banks has fallen from about 14,000 in 1980 to fewer than 7,000 today.³ Even as new banks have entered the industry, there have been over 12,000 bank mergers since 1980, and today, each of the three largest bank holding com-

panies (BHCs)—Bank of America, JPMorgan Chase and Citigroup—has over \$2 trillion in assets. Size is not the only indicator of systemic importance: Some institutions are small but important because of interconnections with other financial firms; others are organizationally very complex.⁴

Some argue that the best way to handle banks that are too big to fail is to break them up.⁵ To evaluate such a solution, it is important to know why banks have gotten so large. Research suggests that some institutions have gotten large, not to game the system, but for reasons of efficiency. The systemic risks posed by large, complex institutions might still outweigh the efficiencies gained by scale, but without estimating these efficiencies, it is impossible to compare costs against benefits. Moreover, the effectiveness of size limits depends on knowing the market pressures on banks that encourage growth. The literature on scale economies in banking, including my own studies, suggests that imposing a strict size limit would have unintended consequences and work against market forces—contrary to both of my guiding principles for regulatory reform.

To my mind, a better solution than legislative limits on bank size is to develop a credible resolution mechanism coupled with other reforms, including revised capital requirements that involve contingent capital and capital charges based on the firm’s contribution to systemic risk, increased disclosures from financial firms, consolidated supervision of large nonbank financial firms, and systemic-risk-focused supervision.

Insights from the literature on scale economies

What has motivated the consolidation of the banking industry?⁶ A growing body of research supports the view that there are significant scale economies in banking. Scale economies are usually measured

with respect to costs and refer to how scale of production (size) is related to costs. A firm is said to be operating with constant returns to scale if, for a given mix of products, a small proportionate increase in all outputs would increase costs by the same proportion. A single-product firm operating with scale economies can lower average cost of production by increasing its scale.

Some cite older research that used data from the 1980s and which did not find scale economies in banking.⁷ The consensus of these earlier studies was that only small banks had the potential for significant scale efficiency gains and the gains were usually small, on the order of 5 percent of costs or less. But more recent studies, using data from the 1990s and 2000s and models of bank production that incorporate risk management aspects of banking, find significant scale economies at even the largest banks in the sample.

Part of the difference appears to reflect improvements in methods used for measuring scale economies,⁸ but it also likely reflects real changes in banking technology, such as computing and telecommunications, and environmental factors, such as a relaxation of governmental restrictions on geographic and product expansion, that have led to a larger efficient scale. The global nature of banking consolidation and increase in scale suggests that U.S. deregulation has not been the only driver. The finding of significant scale economies at banks that are large, but not considered too big to fail, suggests that policy toward the largest institutions is not the only factor.

By their nature, the empirical studies on scale economies derive estimates based on a sample. Constructing samples to include banks that use similar production techniques is important for deriving sound estimates. Newer statistical techniques can overcome some of the drawbacks of earlier studies by fitting the data at the more extreme parts of the sample and not just the sample's average bank. However, only a few existing studies use the most recent data, and bank size has increased significantly over the past 10 years. So, further work needs to be done. Also, the typical estimation techniques do not address whether any particular bank is operating efficiently; other techniques, such as case studies, are more applicable for this type of question. Still, even with these caveats, the studies of scale economies are persuasive that the efficient scale of commercial banking has risen over the past 20 years.

Results of some of the studies

Berger and Mester (1997) estimated the efficiency of almost 6,000 U.S. commercial banks in continuous existence, with complete and accurate data, from 1990 to 1995, and found that about 20 percent of banking costs were lost due to scale inefficiencies, similar to estimates of the loss due to so-called X-inefficiencies (or waste). In every bank size class from less than \$50 million in assets to well over \$10 billion, we found scale economies for more than 90 percent of firms in the size class. In each class, the typical bank would have to be two to three times larger to maximize scale efficiency for its product mix and input prices.⁹ We also found that a simple measure, costs per dollar of gross total assets, displayed scale economies up to \$25 billion in assets, but we concluded that “serious estimates of scale economies for U.S. banks over \$25 billion will likely have to wait for the consolidation of the industry to create enough of these large banks to yield reasonable estimates.” That time has come.

At its heart, banking is about handling risk, and the amount of risk to take on is a management choice. The standard analysis used in earlier studies might not have detected scale economies that actually exist because standard analysis does not account for the risk or capital structure that a bank chooses. A series of papers incorporate managerial preferences over the risk-return trade-off into models of bank production.¹⁰ These studies find that risk management and revenue effects are, indeed, correlated with bank size.

There are two opposing effects on the costs of risk management as banks grow in size. Larger scale may mean better diversification, which could reduce liquidity risk and credit risk. So, there is a diversification effect: Larger scale can lead to reduced marginal cost of risk-taking and reduced marginal cost of risk management, all else equal.

But all else is not necessarily equal because risk-taking is endogenous—a management choice. If banks respond to the lower cost of risk management by taking on more risk in return for greater profits, then we would see another effect of increased scale of operations—a risk-taking effect, which can raise costs, all else equal, if banks have to spend more to manage increased risk or more time dealing with nonperforming assets. Therefore, unless risk is incorporated into the analysis, the increase in costs due to increased risk-taking may

mask scale economies due to diversification.

Hughes, Mester and Moon (2001) found constant returns to scale in a sample of large BHCs using data from 1994 when we used the standard cost-function model from the earlier literature. However, using our more general model incorporating bank managers' preferences about risk and capital structure, we found that BHCs of all sizes were operating with significant returns to scale.¹¹ We also found that large BHCs were operating with less capital than would have minimized their costs and that small banks were operating at more than the cost-minimizing level of capital. And we found evidence of both a diversification effect and a risk-taking effect. Better diversification is associated with larger-scale economies, and increased risk-taking is associated with smaller-scale economies.¹² So the results support the conclusion that scale economies exist, but the usual method cannot find them because it ignores the fact that banks choose their level of risk and their capital structure. Larger scale means lower cost per unit of risk—a scale economy—but it also means banks have the capacity to take on more risk.

Studies that use more recent data are scarce, but those that do exist find significant scale economies in U.S. banking. Using a large data set covering all U.S. commercial banks from 1984 to 2006, Wheelock and Wilson (2009) find that banks had increasing returns to scale throughout the entire distribution of banks—even in 2006, when the largest banks had nearly \$1 trillion in assets. They conclude that “industry consolidation has been driven, at least in part, by scale economies” and that this would imply some cost to limiting bank size. Feng and Serletis (2010), using data from 2000 to 2005 on 293 U.S. banks with over \$1 billion in assets, also find scale economies at the largest banks.

Note that none of the research suggests that regulators should stop considering market power when deciding whether to approve a merger. Indeed, the results are based on banks operating under current regulations and Justice Department guidelines. Nor does the literature suggest that all consolidation and growth is beneficial for society. Too-big-to-fail considerations may be a source of some gains—although not the entire source, since scale economies have been found at banks smaller than those most consider to be too big to fail. Also, other research indicates that managerial entrenchment—that is, the ability of managers to resist market discipline—can lead to inefficient consolidation strategies.¹³

Implications for financial reform

Significant scale economies in banking suggest that economic forces have been an important driver of banks' increasing size. This does not mean that the benefits necessarily outweigh the potential costs that larger size may impose on the financial system and broader economy if size is accompanied by higher risk of systemic problems. But if policymakers do conclude that the costs of size outweigh the benefits, the existence of scale economies suggests that a strict size limit on banks is not likely to be an effective solution. Such limits work against market forces and do not align incentives. Given the potential benefits of size, strict limits would create incentives for firms to avoid these restrictions, and could thereby push risk-taking outside of the regulated financial sector, without necessarily reducing systemic risk.

A better tack would be to increase the costs of becoming too complex or too large commensurate with the risks that these types of institutions impose, for example, imposing a capital charge for contribution to systemic risk, while at the same time trying to close the gaps in supervision. Better understanding of the incentives that financial firms have to avoid supervision and regulation and a focus on macro-prudential supervision of the financial system as a whole will be beneficial in helping to foster financial stability. **R**

Endnotes

¹ The views expressed here are those of the author and do not necessarily represent those of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. An expanded version of this article can be found at <http://www.philadelphiafed.org/research-and-data/economists/mester/>.

² This is not just a theory. Empirical research by Barth, Caprio and Levine (2006) supports this view. They study banking regulatory structures in more than 150 countries and find that transparency and public accountability lead to better banking sector performance than reliance on supervisory discretion.

³ See the Federal Deposit Insurance Corp.'s *Historical Statistics on Banking*, at <http://www2.fdic.gov/hsob/hsobRpt.asp>.

⁴ Rajan (2009) discusses factors other than size that are related to systemic importance.

⁵ See, for example, Johnson and Kwak (2010).

⁶ Mester (2008) provides an overview of methods of measuring productive efficiency in banking and a review of the literature.

⁷ See, for example, Greenspan (2010), p. 32: “For years the Federal Reserve had been concerned about the ever larger size of our financial institutions. Federal Reserve research had been unable to find economies of scale in banking beyond a modest-sized institution.”

⁸ These improvements include using more flexible functional forms to capture the relationship between costs, input prices and output levels; taking into account the bank’s risk and financial capital structure in empirical models; and incorporating banks’ off-balance-sheet activities.

⁹ That both small and large banks operate below efficient scale is not a contradiction; each bank’s level of scale economies is measured based on its own product mix and input prices. Small and large banks choose different product mixes, each suitable to its own scale of operations (see Berger and Mester, 1997). We grouped banks with assets over \$10 billion into a single class because there were too few banks to form credible size classes within this largest category.

¹⁰ See Hughes, Mester and Moon (2001); Hughes, Lang, Mester and Moon (1996, 1999); and Hughes, Lang, Mester, Moon and Pagano (2003). Also, see the summaries in Mester (2008) and Hughes and Mester (2010).

¹¹ Hughes and I are currently working on a study using data from 2007 and 2008.

¹² Diversification referred to the degree of macroeconomic diversification in a BHC’s geographic scope of operations. It was measured by the correlation in unemployment rates over states in which a BHC operates.

¹³ See Hughes, Lang, Mester, Moon and Pagano (2003).

References

Barth, James R., Gerard Caprio, Jr., and Ross Levine. 2006. *Rethinking Bank Regulation: Till Angels Govern*. New York and Cambridge: Cambridge University Press.

Berger, Allen N., and Loretta J. Mester. 1997. “Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?” *Journal of Banking and Finance* 21(July), pp. 895–947.

Feng, Guohua, and Apostolos Serletis. 2010. “Efficiency, Technical Change, and Returns to Scale in Large U.S. Banks: Panel Data Evidence from an Output Distance Function Satisfying Theoretical Regularity.” *Journal of Banking and Finance* 34(1), pp. 127–38.

Greenspan, Alan. 2010. “The Crisis.” Manuscript.

Hughes, Joseph P., William Lang, Loretta J. Mester and Choon-Geol Moon. 1999. “The Dollars and Sense of Bank

Consolidation.” *Journal of Banking and Finance* 23(2/4), pp. 291–324.

Hughes, Joseph P., William Lang, Loretta J. Mester and Choon-Geol Moon. 1996. “Efficient Banking Under Interstate Branching.” *Journal of Money, Credit, and Banking* 28(4), pp. 1045–71.

Hughes, Joseph P., William Lang, Loretta J. Mester, Choon-Geol Moon and Michael Pagano. 2003. “Do Bankers Sacrifice Value to Build Empires? Managerial Incentives, Industry Consolidation, and Financial Performance.” *Journal of Banking and Finance* 27(3), pp. 417–47.

Hughes, Joseph P., and Loretta J. Mester. 2010. “Efficiency in Banking: Theory and Evidence,” in *Oxford Handbook of Banking*, Allen N. Berger, Philip Molyneux, and John O. S. Wilson, eds. Oxford: Oxford University Press, pp. 463–85.

Hughes, Joseph P., Loretta J. Mester and Choon-Geol Moon. 2001. “Are Scale Economies in Banking Elusive or Illusive? Evidence Obtained by Incorporating Capital Structure and Risk-Taking into Models of Bank Production.” *Journal of Banking and Finance* 25(12), pp. 2169–2208.

Johnson, Simon, and James Kwak. 2010. *13 Bankers: The Wall Street Takeover and the Next Financial Meltdown*. New York: Pantheon Books.

Mester, Loretta J. 2008. “Optimal Industrial Structure in Banking,” in *Handbook of Financial Intermediation*, Arnoud Boot and Anjan Thakor, eds. Amsterdam: North-Holland, pp. 133–62.

Rajan, Raghuram. 2009. “Too Systemic to Fail: Consequences, Causes, and Potential Remedies.” Testimony before the Senate Banking Committee, U.S. Senate, May 6.

Wheelock, David C., and Paul W. Wilson. 2009. “Are U.S. Banks too Large?” Working Paper 2009-054B, Federal Reserve Bank of St. Louis. Revised December 2009.