
THE UNITED STATES OF AMERICA
E13806638L
Alan E. Cober

Scale Economies in Banking

Economists have studied the presence of economies of scale in banks for some time. But until recently, this body of research has remained the province of the cognoscenti. The financial crisis and resulting efforts to reform financial regulation have given the topic increased attention among economists and others. Arguing that some banks have grown too big and that size brings with it substantial costs to society—including government bailouts—many prominent observers have advocated breaking up the largest banks. These breakup proponents contend that the economic literature does not find that "large" means more efficient for banks. In other words, they argue that the research shows that in the financial industry as a whole, significant economies of scale do not exist.

Perspective on the bank economies-of-scale literature, in general—and its use in recent regulatory reform, in particular—would therefore be of great value, so I am very pleased that two experts in this area agreed to present their views in a *Region* "symposium." Loretta Mester of the Federal Reserve Bank of Philadelphia and the Wharton School at the University of Pennsylvania discusses the most recent findings from the literature. Robert DeYoung of the University of Kansas, a permanent visiting scholar at the Federal Reserve Bank of Kansas City, highlights some specific limitations on what economists and policymakers can actually know about economies of scale from the current literature. In my essay, I've sought to provide additional context to the overall debate and for the other two essays.

Mester, DeYoung and I share two common, overriding conclusions: First, there remain important unanswered questions about economies of scale in banking. Research that provides answers to such questions will have a high return. Second, even if research shows the presence of economies of scale for large banks, government could potentially improve outcomes by limiting the activities, and size, of these firms.

—Ron Feldman

Size and Regulatory Reform in Finance: Important but Difficult Questions

Ron J. Feldman
Senior Vice President

Many developed countries have experienced financial crises from 2007 to the present, and their governments have responded by, among other steps, protecting creditors of banks from loss. Creditors of large banks have been among the most prominent recipients of government support. Policymakers argue that protecting large bank creditors limited the reduction in economic output that would have resulted otherwise; losses from large banks would have "spilled over" to the broader economy had bailouts not occurred.

Reforms aimed at reducing the likelihood that creditors will receive future bailouts—that is, addressing the too-big-to-fail (TBTF) problem—naturally look to bank size as a potential culprit. Proposed and enacted reforms include putting size caps on banks, limiting bank ability to engage in specific activities, subjecting bank mergers and acquisitions to additional scrutiny and requiring government to proactively break up select banks. Are such reforms good ideas? I am skeptical that reforms focused on size per se will achieve their stated purpose of addressing TBTF; I have more confidence in reforms that identify and address features that produce spillovers in the first place.[1]

Moreover, even if they could address TBTF, reforms that take aim at bank size directly might be bad policy because their costs could exceed their benefits. The size of banks might be positively related to other benefits—that is, big banks could offer cost advantages that would ultimately benefit society. In particular, some banking production processes might benefit from *economies of scale,* wherein the average total cost declines as the quantity of output increases.

Supporters of size-focused reforms generally dismiss the potential for economies of scale in finance. They point to an economics literature that has found scale economies only at firms much smaller than those at the epicenter of the financial crisis. I am sympathetic to this tactic. Indeed, I have used it myself in the past![2] But more recently, I have become dubious of this response, for three reasons.

First, some of the recent econometric work on economies of scale for banking finds such benefits at *all* sizes of banks. Loretta Mester nicely summarizes this extensive research in her *Region* essay. From her review, it's clear that blanket assertions that the "literature" supports one position or another are hard to justify.

Second, and more importantly, we may simply not yet know very much about the presence of scale economies for today's unprecedentedly large banks. Robert DeYoung makes this point in his essay. He argues that the unique nature of today's large banks makes it difficult to apply statistical techniques to historical data to divine the extent of scale economies.

And the limits of our knowledge may go still deeper.

In the first place, it is not entirely clear why the financial sector grew as large as it did in recent years.[3] Banks contribute to economic output through intermediation—that is, by taking in cash from savers and using it to finance projects of households and firms. Banks have performed this economically useful function in many countries, for hundreds of years. Such widespread persistence suggests that banks are particularly adept intermediaries, relative to alternatives.

But value-added intermediation does not justify an infinitely large banking sector. There are reasons to think the sector can be too big in the sense that too many of society's resources are allocated to it.[4] Perceptions by creditors of banks that the government will protect them can lead the sector to grow inefficiently large as TBTF guarantees attract excessive funding to banks. These creditors understand that their bank investments are implicitly subsidized by the assurance of government bailouts should the bank begin to fail.

The market share of banks relative to alternatives like capital markets also varies a great deal over time and place, suggesting that the advantage of banks is not absolute. But this does not mean that alternative markets or institutions could provide intermediation without the potential TBTF downside of banks: The ability of markets and nonbank financial firms to generate potential systemic risk has been clearly demonstrated by the recent financial crisis.

In sum, we do not know if society should continue to rely on banks as much as it does given the potential cost and the alternatives available. This question deserves deep consideration. Calling for more research is a cliché. In this case, the cliché is

apt. Research to better understand the optimal size of the banking sector *could* have high returns.

I emphasize "could" because—and this is a second point about the limits of our knowledge on scale economies—analysts face real challenges in measuring the "output" of banks. Economies of scale relate production size to cost. But what exactly do banks produce? Loans? Deposits? "Liquidity"? Economies-of-scale analysis requires cross-firm comparison. Making sure the comparison is apples to apples is a tremendous challenge. All deposits are not made alike, let alone loans and other products banks offer. Economists working in this area know these and other methodological hurdles well and seek to address them, but the barriers are inherently steep.

Third, the debate about TBTF and scale economies presents the two in contradiction, when in fact they may complement one another. Some activities of a bank—for instance, bank production that relies heavily on automation—may both benefit from scale economies *and* enhance that bank's TBTF status. Banks have automated some types of lending, such as certain credit card and mortgage lending, to a significant degree. Processing of payments, trust and custody services, and provision of Treasury services to firms also depend heavily on automated systems, as would certain types of asset management.

These bank services and products require large investments in automated systems. Once the bank incurs the fixed costs of the systems, it can drive down its total average costs by increasing the volume of goods and services produced. Such automation-dependent products and services can generate a material portion of the revenue banks earn. A superficial guesstimate puts the annual revenue from economies-of-scale services at around 30 percent of the total for one of the largest bank holding companies in the United States.[5]

Many of these automation-based services also enhance TBTF status. Payments processing offers an obvious example. If another bank could not quickly take over or substitute for an important, failing bank provider of payments, important capital markets may not function effectively and some commercial firm payments—perhaps even payroll—would not go through. Even the threat of a payments collapse would lead policymakers to seriously consider all available means to keep the payment train running. Greater scale activity, therefore, could come with higher TBTF cost. The presence of economies of scale, from this perspective, suggests that policymakers sharpen their focus on fixing TBTF. More research on the relationship between larger scale and a more severe TBTF problem therefore seems necessary.

Some bottom lines

Smart people seeking to reduce TBTF have justified policies that would make large banks smaller in part on the basis of published research that does not find significant economies of scale in the financial industry. There are (at least) two reasons that conclusion may not hold. It may not reflect the current state of the literature and, more importantly, it may overstate what we actually do know about such scale economies. Indeed, it may be that banks become more TBTF precisely because they are taking advantage of significant scale economies. More generally, policymakers should focus on addressing the potential for spillovers from failing financial institutions even if scale economies exist. R

Endnotes

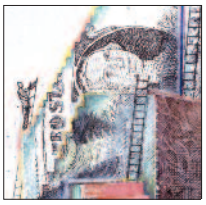
[1] Feldman, Ron J. 2010. "Forcing Financial Institution Change Through Credible Recovery/Resolution Plans." Economic Policy Paper 10-2, Federal Reserve Bank of Minneapolis; Stern, Gary H., and Ron J. Feldman. 2009. "Addressing TBTF by Shrinking Financial Institutions." *Region* 23 (June), Federal Reserve Bank of Minneapolis.

[2] Stern, Gary H., and Ron J. Feldman. 2004. *Too Big To Fail*. Washington, D.C.: Brookings Institution, p. 66.

[3] Philippon, Thomas. 2008. "The Evolution of the U.S. Financial Industry from 1860 to 2007." http://pages.stern.nyu.edu/~tphilipp/papers/finsize.pdf.

[4] There are additional reasons that some observers may view the banking sector as being too big. Some researchers have argued that consumers make systematic errors in their purchases of financial products, including some offered by banks. These errors could lead consumers to consume too many financial products or perhaps the "wrong" financial products. Other observers consider certain activities—such as the trading of financial assets, which sometimes is conducted within a banking organization—as inherently "wasteful." For an example of the first point, see the discussion in John Y. Campbell et al. in "The Regulation of Consumer Financial Products," Social Science Research Network, July 27, 2010, online at papers.ssrn. com. For an example of the second point, see the discussion of a financial transaction tax in "A Fair and Substantial Contribution by the Financial Sector," June 2010, online at imf.org.

[5] To make this extremely rough estimate, we review the fourth quarter 2009 earnings release financial supplement of JPMorgan Chase & Co. In the spirit of Lawrence Radecki (FRBNY *Economic Policy Review,* July 1999), we make the estimate by identifying certain business lines as benefiting from scale and then tallying financial data for these business lines. In particular, we assume that mortgage and credit card lending benefits from scale as do asset management and principal transactions (which include trading activities, among others). The net revenue in 2009 from these operations is $33 billion out of a total of $100 billion. We provide this crude estimate primarily to encourage interested parties to more seriously review bank-specific data to determine the potential importance of scale.



Scale Economies in Banking and Financial Regulatory Reform[1]

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 companies (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., Willim 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.

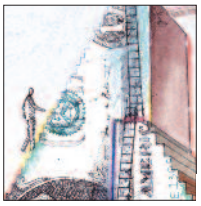
Hughes, Joseph P., Loretta J. Mester and Choon-Geol Moon. 2001. "Are Scale Economies in Banking Elusive or Illusive? Evidence Obtained by Incorporating Capital Structure and Risk-Taking into Models of Bank Production." *Journal of Banking and Finance* 25(12), pp. 2169–2208.

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.



Scale Economies Are a Distraction

The real issue for policy is credible resolution of failing financial firms, not bank size

Robert DeYoung
Federal Reserve Bank of Kansas City
and University of Kansas

A small cadre of banking economists (including, for a time, me) has studied banking companies for nearly half a century in an effort to answer the following question: Can banks become more efficient by growing larger? Or, in the technical vernacular, do banks exhibit *scale economies*? This question has garnered fresh attention today as policymakers consider steps to regulate bank size in light of too-big-to-fail concerns.

Possible scale economies in the banking industry were also a crucial question for bank regulatory policy during the 1980s and 1990s. Existing regulations kept banks small by prohibiting their expansion across state lines; bankers argued that these rules made the U.S. banking system inefficient. Removing these constraints, they said, would enable them to expand their geographic footprints and capture scale economies. And because banking services are sold in competitive markets, much of the resulting cost savings would be passed along to customers and not simply accrue to bank shareholders.

The question of scale economies was important for banks of all sizes. If two small banks from neighboring states merged, would running the resulting medium-sized bank be cheaper than running the two small banks separately? What if two medium-sized banks merged to create a regional bank? Or if two regional banks merged to create a bank with national presence?

According to the earliest statistical studies, scale economies "ran out" once a bank had accumulated assets of $100 million or $200 million—that is, only small banks could hope to capture scale economies by growing larger. But as my research colleagues developed new and better analytical tools, their conclusions evolved. Subsequent studies found available scale economies up to $500 million in assets … then $1 billion … then $10 billion to $25 billion—that is, all but a handful of U.S. banks at the time had access to scale economies. By the mid-1990s, some of the more innovative studies were reporting that, under certain circumstances, even the largest banks had access to scale economies.[1]

In retrospect, those scale economy studies were the right tool for the job. They provided objective evidence on an argument being made by the (perhaps less than objective) financial services industry. In a significant way, those studies helped pave the way for deregulation and the mix of local, regional and national banks in existence today.

Scale economies redux

The last of the major restrictions on banking geography were removed in 1997 when the Riegle-Neal Act was implemented. In the wave of industry consolidation that ensued, banks of all sizes grew larger by acquiring banks in other states.

At the upper end, the merger wave created banking companies far larger than the banks examined in the scale economy studies of the 1980s and 1990s. For example, today the three largest U.S. banking firms (Bank of America, JPMorgan Chase and CitiGroup) all exceed $2 trillion in assets, while the three next largest (Wells Fargo, Goldman Sachs and Morgan Stanley) all have assets in excess of $800 billion, well above the range covered by academic researchers.

In 2008 and 2009, some of these banking giants suffered huge financial losses that, by virtue of their size alone, threatened the stability of financial markets and the macroeconomy. Government policymakers judged that the risks of allowing those firms to fail were too great; famously, financially troubled banking firms received hundreds of billions of dollars in capital injections and other forms of taxpayer-backed bailouts.

Preventing such an episode from happening again was the focus of long congressional debates this year over legislation to reregulate financial institutions. However, the new law that emerged

leaves important questions related to bank size unanswered: Should the public continue to live with these large banks and the risks they impose? Should regulators break up these firms? Or should policy give these firms incentives to downsize, such as imposing size-based taxes or higher capital requirements?

Clearly, understanding the existence and/or limits of bank scale economies is once again important for forming public policy. But the nature of this inquiry is different from the deregulatory questions of the 1980s and 1990s. First, policymakers and researchers are now interested only in scale economies at the very largest banks, not at banks of all sizes. Second, policymakers now need to know whether any resulting efficiencies are substantial enough to justify living with the social costs and macroeconomic risks posed by these newly enormous firms.

Despite the hard and often ingenious work of my colleagues in the bank scale economy field, I am not optimistic that this line of research will generate the answers needed this time around. Why not? The standard approaches to measuring scale economies are the least accurate for precisely those firms most relevant to the question at hand: the very largest banking companies.

The wrong tool for the job

It is well-known that the statistical techniques employed to measure scale economies in any industry deliver the most accurate estimates for "average" companies in that industry; for firms that are substantially smaller or larger than average, estimates grow increasingly less precise. This characteristic is especially problematic for the banking industry, due to the drastically skewed size distribution of its firms. As of March 2010, the three largest banking companies (mentioned above) each had assets of over $2 trillion, *10 times* larger than the 13th-largest banking company, Bank of New York Mellon, with assets of $220 billion. They were *100 times large*r than the 43rd-largest bank, BOK Financial of Tulsa, Okla., with assets of $23 billion. Because of these dramatic size differences, statistical estimates of scale economies among large banks can be quite sensitive to the good or bad financial fortunes of just one or two of these largest banks.

A second problem arises because the largest banks operate quite differently than small and medium-sized banks; that is, they differ in kind, not just size. But because most of the available data come from the thousands of small and medium banks, bank scale economy models are based on the business processes most often used by these banks. This segment of the industry relies predominantly on traditional banking approaches: holding illiquid loans, issuing liquid deposits to finance those loans and earning profits chiefly from the resulting interest margin. But the very largest banking companies produce financial services quite differently. They rely less on deposits and more on short-term market financing, they sell many of their loans rather than hold them, and they earn a substantial portion of their profits from customer fees rather than interest margins. Using models built around smaller bank production processes to describe the relative efficiency of large banking companies can be misleading.

These methodological deficiencies did not prevent scale economy studies from usefully informing the deregulation debate of the 1980s and 1990s. Geographic deregulation was relevant for banks of all sizes and, at that time, bank production processes were still pretty similar for large and small banks. But these issues may be debilitating in today's debate over reregulating the largest banking companies—while scale economies *might* exist for these banking giants, we cannot be sure because measuring these phenomena stretches our analytic tools to, and perhaps beyond, their limits.

What about market forces?

Perhaps there is a simpler way. Rather than estimating complex models of bank scale economies, could we simply depend on the market to reveal the best size for banks?

The argument goes like this: The fact that banks have grown increasingly large over time is *prima facie* evidence that scale economies exist for even the largest banks. If this were not the case, managers of large banks would be operating inefficiently large firms, and their ill-served shareholders would attribute lower profits to *dis*economies of scale and sell their shares. Investors would purchase, pull apart and reallocate the assets of these firms.[2] Thus, market discipline would ensure that banks would exhibit the most profitable range of sizes and other attributes.

While I generally embrace this line of reasoning, the argument fails for the very largest banking com-

panies in the United States today. Even if these banks are too large to operate efficiently, shareholders are unlikely to recognize or act on this, because the performance-detracting effects of scale diseconomies are masked by the performance-enhancing effects of the too-big-too-fail subsidies enjoyed by these banks. Given the government bailouts of 2008 and 2009, there is no longer any doubt that the largest U.S. financial companies are considered too big to fail. Because these firms can perform poorly and still remain in business, shareholders and creditors benefit from upside success without suffering the full downside losses, which gives the largest banking companies a cost-of-capital advantage over their smaller rivals. In other words, there may be the appearance of scale economies for these firms where none really exists.

Focus on resolution policy, not bank size

If we cannot confidently measure scale economies at the very largest banking companies—and indeed, although researchers have attempted methodological "fixes" of the deficiencies I've mentioned above, I am not sure that we can—then are we forced to make uninformed regulatory policies for these firms? Must we make decisions about whether to break up, downsize or somehow limit the growth of these institutions without reasonable certainty as to the consequences of such actions for the future efficiency of the banking sector?

My sense is that the question of scale economies in banking, while of real interest, is something of a distraction to the primary issue. The chief concern should be not how big banks must be to achieve optimal efficiency, but rather, how policymakers can establish a credible strategy for resolving banks when they fail—regardless of their size, complexity and inter-connectedness. The public needs policies and policymakers that impose harsh discipline on the managers, shareholders and junior debt holders of large failed banks—while simultaneously using bridge banks, other available resolution techniques and expanded resolution authority to preserve the liquidity of borrowers, depositors and other counterparties of these banks.

Of course, this is a tall order. But the current inability to do this is the root cause of the too-big-to-fail problem often attributed to bank size. And by addressing this root cause—rather than placing regulatory limits on bank assets or some other measure of size, an ad hoc policy that will surely result in unintended consequences—we will generate a number of benefits. Chief among them: The primary justification for too-big-to-fail subsidies would disappear. Large banks might continue to pose a problem for competitive efficiency (a concern of antitrust policy), but no longer for macroeconomic stability. And we could then rely on the marketplace—no longer handicapped by poorly designed policy—to reveal the optimal size for banks. R

Endnotes

[1] An article by Allen Berger, Rebecca Demsetz and Philip Strahan in the February 1999 *Journal of Banking and Finance* discusses this literature in more detail (see pages 157-60). While the advancing research has found increasing access to scale economies for banks, no similar consensus has emerged regarding the dollar magnitudes of these savings or whether managers running large banks are able to fully exploit the potential for savings.

[2] Because changes in ownership of banks require regulatory approvals, this "market for corporate control" mechanism would likely work more slowly in the banking industry than in other industries.