Hughes, Joseph P.

**Working Paper**

# The elusive scale economies of the largest banks and their implications for global competitiveness

Working Paper, No. 2011-34

**Provided in Cooperation with:**
Department of Economics, Rutgers University

*Suggested Citation:* Hughes, Joseph P. (2011) : The elusive scale economies of the largest banks and their implications for global competitiveness, Working Paper, No. 2011-34, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:
https://hdl.handle.net/10419/59504

# THE ELUSIVE SCALE ECONOMIES OF THE LARGEST BANKS AND THEIR IMPLICATIONS FOR GLOBAL COMPETITIVENESS*

**Joseph P. Hughes**

**Department of Economics**
**Rutgers University**

**October 2011**
**Revised December 6, 2011**

**Fourteenth Annual International Banking Conference**
**Federal Reserve Bank of Chicago**
**in conjunction with the European Central Bank**
**November 10-11, 2011**

## Abstract

In the wake of the financial crisis that began in 2007, policy makers have focused again on the largest financial firms to consider the association of their size with systemic risk. An equally important question examines whether their size benefits the economy. In particular, is the size of our largest financial institutions the result of technological cost advantages that improve the efficiency of their capital allocation and liquidity and enhance their international competitiveness? Or is it the result, not of technological cost advantages, but of safety-net subsidies that confer too-big-to-fail cost advantages and foster moral hazard in investment decisions.

This paper reviews the evidence of large scale economies that increase with size and considers the credibility of this evidence by examining details of how scale economies are measured and why evidence of scale economies eludes many investigations. A method of estimating scale economies developed by Hughes, Lang, Mester, and Moon (1996) distinguishes the underlying scale effects on cost from the effects on costs of size-related changes in risk-taking, which can obscure technological cost advantages, such as those due to better diversification. It reviews evidence that technology, not too-big-to-fail subsidies, accounts for the cost advantage of the largest financial institutions. Finally, it considers the implications of scale economies for scaling back the operations of the largest financial institutions and for the global competitiveness of smaller institutions.

*The author is Professor of Economics at Rutgers University, New Brunswick, New Jersey.

**Introduction**

In the wake of the financial crisis that began in 2007, policy makers have focused again on the largest financial firms to consider the association of their size with systemic risk. An equally important question examines whether their size benefits the economy. In particular, is the size of the largest financial institutions the result of technological cost advantages that improve the efficiency of their capital allocation and liquidity and enhance their international competitiveness? Or is their large size the result, not of technological cost advantages, but of safety-net subsidies that confer too-big-to-fail cost advantages and foster moral hazard in investment decisions? The answers to these questions help focus and balance the debate on regulating the size of the largest financial institutions to manage systemic risk.

In a recent speech at the Conference on the Regulation of Systemic Risk at Federal Reserve Board in Washington, Federal Reserve Governor Daniel Tarullo noted the relative dearth of empirical work on the degree of scale and scope economies in large financial firms: "Generally, though, even where intuition suggests economies in some other areas – such as the breadth of securities distribution networks and the ability to provide all forms of financing in significant amounts – evidence for the existence of such economies is limited and mixed. Moreover, even where significant scale is necessary to achieve certain economies, an important question will be what the minimum efficient scale – or, perhaps more realistically, the minimum feasible scale – actually is. It is possible that a firm would need to be quite large and diversified to achieve these economies, but still not as large and diversified as some of today's firms have become."

Governor Tarullo poses two fundamental questions. Why is the evidence of scale economies limited and mixed – are they illusive or elusive? And, if such economies exist, is it

possible that they can be achieved by smaller institutions than the largest we observe today? My research in collaboration with Loretta Mester, Choon-Geol Moon, and William Lang has sought to answer these two questions.[1] We have consistently found evidence of large scale economies that increase with size – the largest financial institutions experience the largest scale economies, and we have shown why this evidence eludes the standard investigation. In recent work Loretta Mester and I have found no evidence that too-big-to-fail subsidies in the cost of funds explain the cost advantages of the largest financial institutions, and we have found that restricting the size of the largest institutions would significantly increase the cost of producing their current menu of financial products and services, an increase which would likely compromise their global competitiveness.[2] The following sections address each of these points. The first section reviews the evidence of large scale economies that increase with size. The second section considers details of how scale economies are measured and why evidence of scale economies eludes many investigations. The third section outlines a method of estimating scale economies that distinguishes the underlying scale effects on cost from the effects on costs of size-related changes in risk-taking. The fourth section describes evidence that technology, not too-big-to-fail subsidies, account for the cost advantage of the largest financial institutions. The fifth section considers the implications of scale economies for scaling back the operations of the largest financial institutions and for the global competitiveness of smaller institutions.

**Evidence of scale economies at the largest financial institutions**

The term, economies of scale, describes how cost varies with a measure of firm size which characterizes the financial products and services produced by the institutions. When cost

[1] Hughes, Lang, Mester, and Moon (1996, 2000), Hughes and Mester (1998), Hughes, Mester, and Moon (2001).
[2] Hughes and Mester (2011)

increases less than proportionately with size, economies of scale prevail. When cost varies in the same proportion as size, constant returns to scale obtains. And, when cost increases more than proportionately with size, diseconomies of scale prevail.

Using a variety of recently developed methods of modeling the relationship of cost to financial outputs, a number of studies have found evidence of economically and statistically significant scale economies even at the largest financial institutions. These studies are described briefly below and in more detail in Hughes and Mester (2010). Hughes, Lang, Mester, and Moon (1996) use 1994 data consisting of 443 highest-level U. S. bank holding companies ranging in size from $33 million to $250 billion in consolidated assets. They find that the cost elasticity, 0.89, of the smallest banks (assets less than $300 million) decreases continuously as bank size increases up to a mean of 0.80 for banks whose assets exceed $50 billion.[3] A value of 0.89 for the cost elasticity implies that a 10 percent increase in output would result in an 8.9 percent increase in cost, and for 0.80, an 8.0 percent increase in cost. The less than proportional response of cost to an increase in output implies that economies of scale characterize technology. This study finds the largest scale economies at the largest financial institutions.

Berger and Mester (1997) use U. S. data on commercial banks during the period 1990-1995 and find that in each size class from banks with assets less than $50 million in assets to those with assets that exceed $10 billion, over 90 percent of banks in each size group experiences scale economies. The typical bank in each size class would need to be two to three times larger to maximize scale efficiency.

Hughes and Mester (1998) examine U. S. commercial banks with assets that exceed $1 billion during the period 1989-1990 and find significant scale economies for banks in all size

[3] The measure of scale economies is often stated as the inverse of the cost elasticity. The cost elasticity 0.89 implies a measure, 1.12, of scale economies. The cost elasticity of 0.80 implies scale economies of 1.25. Thus, scale economies increase with size as the cost elasticity decreases.

groups. For the smallest banks with assets less than $1.8 billion, a 10 percent increase in output implied a 9.4 percent increase in cost while, for the largest banks with assets between $7 billion and $74 billion, an 8.7 percent increase in cost. Hughes, Lang, Mester, and Moon (2000) consider the same sample but employ a more innovative technique to model banking technology and find that a 10 percent increase in output implies a 9.1 percent increase in cost for the smallest banks and an 8.3 percent increase for the largest banks.

DeYoung, Hughes, and Moon (2001) use the method of Hughes, Lang, Mester, and Moon (1996) to study of the relationship of Camel ratings to efficiency. Their sample consists of 356 national banks in the U. S. in 1994. Total assets ranges from $83.6 billion to $120 billion. In results not reported in the paper, they find a mean cost elasticity for the full sample of 0.893 which ranges from 0.926 for the smallest quartile to 0.826 for the largest quartile.

Bossone and Lee (2004) also apply the method of Hughes, Lang, Mester, and Moon (1996) to data on 875 commercial banks from 75 countries. They find significant scale economies that increase with the size of the country's financial system. They term these economies "systemic scale economies."

Wheelock and Wilson (2010) also use another innovative approach to measure scale economies at all U.S. banks over the period 1986- 2004. They find that over the entire period most banks experienced scale economies.

Feng and Serletis (2010) examine U. S. banks with assets greater than $1 billion over the period 2000-2005 and find scale economies at all sizes. On average, a 10 percent increase in output is associated with a 9.5 percent increase in cost.

Hughes and Mester (2011) employ 2007 data on 842 highest-level U. S. bank holding companies to estimate the production model developed by Hughes, Lang, Mester, and Moon

(1996). The companies in their sample range in size from $72 million in total assets to $2.2 trillion. For institutions with less than $0.8 billion in total assets, they find that a 10 percent increase in output is associated with an 8.8 percent increase in cost. For larger institutions with assets between $50 and $100 billion, the increase in cost is 8.1 percent. For the 17 institutions whose assets exceed $100 billion – a category they term "too-big-to-fail" – the increase in cost is 7.5 percent.[4] They also test and reject the proposition that too-big-to-fail subsidies generate these scale economies of the largest institutions.

**How are economies of scale typically measured and why do they elude many investigations?**

The simplicity of the definition of scale economies – how cost varies with output -- obscures some important complications. First, how is cost defined? Second, how is output or size defined? Third, how is the evidence of their association uncovered?

Cost is usually defined as the sum of interest and noninterest expenses, which represent the costs associated with the factors of production – borrowed funds, labor, and physical capital. Hughes and Mester (2011) define six inputs: labor, physical capital, uninsured time deposits, all other deposits, all other borrowed funds (foreign deposits, federal funds purchased, reverse repos, trading account liabilities, mandatory convertible securities, mortgage indebtedness, commercial paper, and all other borrowed funds), and equity capital. The measure of cost sums the expenditures on all inputs for which a price can be computed. The cost function includes these input prices to account for differences in prices among financial institutions. In the case of

[4] Brewer and Jagtiani (2009) list three definitions of too big to fail: (1) institutions whose book-value of consolidated assets exceeds $100 billion, (2) one of the 11 largest banks, a definition offered by the Controller of the Currency in 1984 (currently the 11th largest has $290 billion in assets), and (3) banks with market value of equity of $20 billion.

too-big-to-fail institutions, any advantage they enjoy in their funding costs is taken into account by the prices of their borrowed funds so that, in principle, the cost playing field is leveled. The expense of equity capital is usually excluded since most financial institutions are not publicly traded and, hence, this expense cannot practically be computed. The theoretically correct step to take in such cases requires the inclusion of the quantity of equity capital in lieu of the required return and cost of capital. Including the quantity of equity permits the calculation of a shadow required return and cost of capital for all financial institutions.[5]

Total assets and total earning assets constitute naïve measures of size. Two problems with these measures require a less aggregated characterization of size. First, off-balance-sheet activities represent an important part of many financial institutions' output. Second, the costs of two banks with the same total assets and off-balance-sheet activities might differ substantially because their asset allocations differ – say, because one allocates a larger proportion to loans and less to liquid assets. Liquid assets are inexpensive to obtain relative to information-intensive loans. Defining disaggregated outputs must balance the benefits of detail against the costs of statistical complexity that result when the number of outputs is increased. This balance involves some degree of aggregation. However, it should contain sufficient detail that the types of output span the differences in investment strategies across institutions of all sizes. Hughes and Mester (2011) define five outputs: liquid assets (including cash, repos, federal funds sold, and interest-bearing deposits due from banks), securities (including U.S. Treasury and U.S. government agency securities as well as mortgage-backed securities), loans, trading assets, and the credit equivalent amount of off-balance-sheet activities. The differences among small and large banks

[5] Loan losses as well as the cost of equity capital are usually excluded from the measure of total costs used to estimate the cost function. Hughes and Mester (2010, 2011) explain the role of the quantities of equity capital and nonperforming loans in the cost function by reference to the standard theoretical result that the quantity of a factor of production can substitute in the cost function for the price of the factor and the associated expense. See also Hughes, Mester, and Moon (2001).

in their mix of these five types of outputs facilitates the statistical estimation of the relationship of cost to output for banks of all sizes when a sufficiently flexible functional form is used to model the relationship.

Building on these ingredients, the standard study employs statistical analysis to estimate the relationship of measured cost to the characterization of input prices and outputs and, if correctly specified, the quantity or equity capital and nonperforming loans. It usually finds evidence of economies of scale at smaller institutions and either constant returns to scale or diseconomies of scale at larger institutions. At least three considerations call these findings of little or no scale economies into question: first, textbook explanations of why larger institutions should enjoy scale economies; second, the historically growing size of the largest financial institutions; and, third, mergers that create large institutions. Of course, these large institutions may have sought their large scale to achieve the status of too big to fail – a potential advantage that might offset any scale diseconomies. Nevertheless, my coauthors and I have consistently found evidence that scale economies at the largest financial institutions are elusive, not illusive.[6]

What makes scale economies so elusive? The standard investigation estimates the relationship of cost to financial outputs using a characterization of cost and output similar to the ones described above. The details of the relationship to be estimated follow from a mathematical model of cost minimization. Given the interest rates of the various sources of funding, the prices of labor and physical capital, and the quantity of equity capital – the inputs in the production process – the cost of producing any given quantities of the outputs at their quality measured by nonperforming loans is assumed to be minimized, and the resulting cost function is fitted to the data statistically with a flexible functional form.

[6] Hughes, Lang, Mester, and Moon (1996, 2000), Hughes and Mester (1998), Hughes, Mester, and Moon (2001).

While this procedure may work well in explaining the costs of many nonfinancial industries, it misses a key ingredient in the production of financial products and services: ***risk***. In particular, the size of the financial institution influences the institution's diversification of liquidity and credit risk. Textbooks assert that larger, better diversified institutions experience relatively lower costs of risk management than smaller, less diversified institutions. Consequently, larger institutions can economize on holdings of liquid assets without increasing their liquidity risk and on equity capital without increasing their insolvency risk. This assertion implies that, other things equal, cost increases less than proportionately with output – that is, better size-related diversification tends to generate scale economies. Hughes, Mester, and Moon (2001) estimate a variant of the standard cost function and find diseconomies of scale at smaller banks and constant returns at the largest banks. As hypothesized, they show that its measure of scale economies is positively correlated with asset size and, more to the point, positively correlated with a measure of the geographic diversification of macroeconomic risk. Why, then, do scale economies elude the standard investigation?

They elude the investigations that ignore risk because other things are not equal – in particular, risk. Consider how better diversification influences the investment decisions of larger institutions. *Is the larger, better diversified institution less risky? Not necessarily – better diversification improves the larger institution's risk-expected-return frontier. It does not imply that the larger institution chooses a risk exposure on the improved frontier which is lower than on the less diversified frontier*. Hughes (1999) describes how the effect of additional risk-taking on cost may obscure inherent scale economies due to better diversification. The better risk-return menu and lower marginal cost of risk management may encourage the larger financial institution to choose a risker investment strategy and, consequently, incur potentially higher costs of risk

management. To the extent that any higher costs occasioned by the additional risk-taking offset or overtake any cost reductions due to better diversification, the naïve estimation of the relationship of cost to output that ignores endogenous risk-taking might find, not that cost increases less than proportionately with output, but that cost increases proportionately or more than proportionately. In such a case, *the scale economies that are occasioned by better diversification are obscured by extra costs of additional risk-taking.* The finding of no economies or even diseconomies of scale at larger institutions is a typical research finding.[7] Hughes, Mester, and Moon (2001) label these two effects of better scale-related diversification the *diversification effect* and the *risk-taking effect*.

Hughes and Mester (2011) illustrate the diversification and risk-taking effects identified in Hughes (1999) and Hughes, Mester, and Moon (2001) in a figure with investment choices on two risk-expected-return frontiers, reproduced here as Figure 1. The inferior frontier reflects the trade-off of a smaller, less diversified scale. Assume the smaller institution chooses a mix of financial products and services, asset quality, and funding strategies that implies the trade-off at point A. Now consider a proportionately larger institution whose scale-related better diversification gives it the improved risk-return frontier. As the scale of the mix of financial products and services is increased, does cost increase less than proportionately, proportionately, or more than proportionately? If the larger institution takes no additional risks compared to the smaller institution at point A, the better diversification reduces its risk and improves its expected return, which is reflected by point B. Consequently, *cost at point B increases less than proportionately compared to cost at point A, which implies economies of scale.* These scale economies reflect the better diversification of the underlying larger scale of operations.

[7] See Greenspan (2010), Financial Oversight Council (2011), and Tarullo (2011).

How might these economies elude detection? Suppose the larger institution takes advantage of the reduced marginal cost of risk management and incurs additional risk for a higher expected return – say point C. At C, the risk of the larger institution equals that of the smaller one, but its expected return is higher. For example, point C might result from a reduction in asset quality or an increase in financial leverage. To the extent that the additional risk-taking results in higher costs than at point B, *cost at point C may increase proportionately or even more than proportionately compared to point A.* If cost increases in proportion with the increase in output from the lower to the higher frontier, the standard estimation of scale economies would obtain constant returns to scale. Thus, the additional risk-taking obscures the inherent scale economies and gives the appearance of constant returns to scale. At D, the larger institution takes more risk than the smaller institution so that the *cost at point D is likely to increase more than proportionately compared to point A as output increases from the lower to the higher frontier, which gives the misleading appearance of diseconomies of scale*.

**How are scale economies measured while accounting for endogenous risk-taking?**

In measuring scale economies, accounting for endogenous risk-taking is essential. Larger institutions generally have the incentive to take more risk – to operate toward point D rather than point B, which tends to obscure their inherent scale economies from the standard analysis. Marcus (1984) has shown that managers of financial institutions face dichotomous risk-taking strategies for maximizing the value of their firms that result from limitations on entry into banking and from safety-net subsidies. Entry restrictions create market power which is especially valuable in markets with high-valued investment opportunities. Managers whose institutions operate in such markets enhance value by pursuing less risky investment strategies to

reduce the risk of financial distress and the potential loss of the valuable charter to operate in these markets. On the other hand, managers whose institutions operate in more competitive markets with lower valued investment opportunities enhance value by pursuing higher risk investment strategies to exploit the mispriced safety net. Marcus (1984) pointed out that these incentives are dichotomous: midrange risk-taking strategies do not maximize value. Hughes, Lang, Moon, and Pagano (1997) provide evidence of these dichotomous investment strategies. Grossman (1992) documents the risk-taking encouraged by mispriced deposit insurance. Keeley (1990) provides evidence that increased competition among U. S. banks in the period 1971-1986 reduced the value of banks' investment opportunities and encouraged additional risk-taking. Tufano (1996) contends that, when there is the potential for financial distress and, by extension, when the mispriced safety net subsidizes risk-taking, managing risk is a risk-neutral strategy to maximize value.

Larger institutions typically exploit the safety-net subsidies while smaller institutions typically pursue a less risky investment strategy to protect their valuable charter (Hughes, Lang, Moon, and Pagano, 1997). This pattern suggests that, to the extent the extra risk-taking of the larger institutions involves more costs that offset the reduction in cost due to their better diversification, the extra costs due to risk-taking may obscure their scale economies when endogenous risk-taking is not taken into account in estimating the cost-output relationship.[8]

The standard analysis of cost assumes that, given the prices of the inputs, the mix of financial products and services is produced with the lowest cost mix of inputs. This mix of inputs includes various funding sources of differing maturities as well as labor and physical

[8] Some additional risk-taking may be less costly. For example, a bank may skimp on the resources it devotes to credit analysis. Its risk and expected return increase due to this particular cost saving. However, if banks generally responded to better scale-related diversification by skimping, the standard cost analysis that does not account for risk should find economies of scale rather than constant returns or diseconomies of scale.

capital. Differences among mixes involving liquidity risk, credit risk, and other risk exposures are not taken into account in this analysis. Note, too, that these different risk exposures involve different expected returns. Thus, when value-maximizing firms choose more costly input mixes because they are managing their risk exposure to achieve higher expected returns, the standard analysis of cost is likely to confuse the extra cost that may result from additional risk-taking with a lack of inherent scale economies and fail to identify the underlying economies.

In a series of papers beginning with Hughes, Lang, Mester, and Moon (1996), the authors propose a cost function that accounts for endogenous risk-taking and, hence, is able to identify the underlying scale economies that may be obscured when larger financial institutions take extra risk to exploit the safety net.[9] They provide an econometric model which estimates managers' rankings of risk and expected return – their choice along their risk-return frontier. Risk is characterized in terms of the production plan, and cost is inferred from the choice of expected return and the underlying production plan. Unlike the standard approach, cost in this model depends, not just on the prices of funding sources, equity, labor, and physical capital, but also on revenue considerations and marginal tax rates. Revenue can drive cost. This model is sufficiently general that it subsumes the standard cost minimization model as a special case. In this special case of cost minimization, revenue, risk, return, and tax rates do not influence production decisions. The authors test the hypothesis that these variables have no influence on production and, in every case, cleanly reject the hypothesis.[10]

They illustrate how underlying scale economies elude the standard analysis by estimating scale economies from their risk-return-driven cost function and from several versions of the

[9] Other papers of these authors which develop the risk-return-driven cost function include Hughes, Lang, Mester, and Moon (2000), Hughes, Mester, and Moon (2001), DeYoung, Hughes, and Moon (2001), Hughes and Mester (2010), and Hughes and Mester (2011).

[10] See Hughes, Lang, Mester, and Moon (1996, 2000); Hughes, Mester, and Moon (2001); DeYoung, Hughes, and Moon (2001).

standard cost function. Table 1 presents their estimates for the 1994 data (Hughes, Mester, and Moon (2001)). In column 1 of Table 1, the cost elasticities are obtained from the estimation of a cost function which omits the cost of equity capital in the measure of total cost and does not control for the amount of equity, which is theoretically required when the expense is not included in total cost. Thus, this cost function, while a commonly used one, is mis-specified. The full sample and size groups larger than $0.3 billion and smaller than $10 billion in consolidated assets show slight scale economies. For the full sample, a 10 percent increase in outputs would imply a 9.885 percent increase in cost. The cost elasticity of institutions between $10 billion and $50 billion, the elasticity 0.9924 is not statistically different from one. Neither is the elasticity, 0.9922, for institutions larger than $50 billion. Thus, the largest U. S. financial institutions in 1994 appear to exhibit constant returns to scale. In column 2, the measured cost includes an estimated expense for equity. The results show that for the full sample and for the size groups up to $50 billion, institutions on average exhibit small diseconomies of scale. For the institutions with more than $50 billion, the mean cost elasticity, 1.0130, is not statistically different from one. Thus, the cost elasticity at the largest banks exhibits constant returns to scale. For both estimations of the standard approach, any underlying economies of scale at the largest financial institutions have eluded measurement.

In column 3 of Table 1, the evidence derived from the risk-return-driven cost function shows a mean cost elasticity of 0.8949 for the smallest group of institutions with consolidated assets less than $0.3 billion – a finding of economies of scale – which decreases continuously to a value of 0.7998 for the largest group with assets exceeding $50 billion. This approach which accounts for risk-expected-return choices in measuring costs uncovers evidence of large scale

economies which increase with the size of the institution. The largest financial institutions experience the largest economies.

How typical is the year 1994? Does the evidence of such scale economies from the risk-return-driven cost function and the lack of it from the standard cost functions depend on this particular year? Hughes, Lang, Mester, and Moon estimated the risk-return-driven cost function for 286 U.S. commercial banks in 1990 whose total assets exceeded $1 billion. The total assets of these banks ranged from $1.025 billion to $69.612 billion. Table 2 reports their mean estimates of cost elasticities from the standard, mis-specified cost function by size groups for these banks. The full sample exhibits a mean cost elasticity of 0.967 while the smallest quartile experiences a mean elasticity of 0.979 which decreases to 0.952 for the largest quartile. Hence, the standard, mis-specified cost function exhibits slight cost economies for banks in all size groups.

On the other hand, the estimation of the risk-return-driven cost function for these banks uncovers large scale economies that increase with banks' total assets. Column (2) shows that a 10 percent increase in the outputs of banks in the smallest quartile would be associated with a 9.08 percent increase in cost while in the largest quartile, an 8.28 percent increase in cost. Thus, banks of all sizes experience on average large economies of scale that increase with size. Column (3) reports the mean cost elasticities for the 1994 data from Table 2 organized by the size grouping of the 1990 data and deflated to constant 1990 dollars. The 10 percent increase in outputs would be associated with a mean 8.95 increase in cost in the smallest asset quartile and a mean 8.24 increase in the largest quartile. The mean cost elasticities in 1990 for commercial banks and in 1994 for bank holding companies are remarkably similar and suggest that banks experience scale economies that increase with their total assets.

Since the period of 1990 to 1994, a number of technological advances, especially advances in information technology, might be expected to augment scale economies. Moreover, financial institutions have grown appreciably in scale. In a recent working paper, Hughes and Mester (2011) report on their investigation of scale economies at 842 top-tier U. S. bank holding companies in 2007. They employ the production model developed by Hughes, Lang, Mester, and Moon (1996). The companies in their sample range in size from $72 million to $2.2 trillion in total consolidated assets. For institutions with less than $0.8 billion in total assets, they find that a 10 percent increase in output is associated with an 8.8 percent increase in cost. For larger institutions with assets between $50 and $100 billion, the increase in cost is 8.1 percent. For the 17 too-big-to-fail institutions whose assets exceed $100 billion the increase in cost is 7.5 percent. They conduct several robustness tests to check, first, that institutions with unusual output allocations and, second, that institutions at the smallest and largest ends of the size distribution do not overly influence the estimation of scale economies. Their findings are robust to these tests.

**Do too-big-to-fail subsidies account for the estimated scale economies at the largest institutions?**

Since larger institutions may enjoy a cost-of-funds advantage due to the safety net and since input prices vary across institutions for a variety of reasons, the standard cost function and the risk-return-driven cost function both attempt to level the playing field across institutions by controlling for the prices of inputs – in particular, the average interest rates on various types of borrowed funds paid by each institution. Thus, if the largest institutions experience a cost-of-funds subsidy due to their too-big-to-fail status, this advantage is taken into account in the

estimation of cost – that is, cost is conditioned on the interest rates paid for borrowed funds to level the price playing field – so that the cost elasticity is estimated *given the prices paid for these funds*. To check that the playing field is indeed leveled, Hughes and Mester (2011) calculate scale economies for the too-big-to-fail institutions by using the prices of borrowed funds paid by the smaller banks and obtain estimates of scale economies that are remarkably similar to those obtained using the largest banks' own prices. Finally, they drop the too-big-to-fail institutions, those whose consolidated assets exceed $100 billion, and re-estimate the production model. The estimated scale economies for these largest financial institutions, predicted out of sample, are essentially the same as those from the estimation that included these institutions.

They conclude that there is no evidence from these tests that any too-big-to-fail cost advantage of the largest financial institutions generates the estimated economies of scale. Thus, technological factors appear responsible.

**What are the implications of the estimated scale economies at the largest institutions for restricting their scale and for their international competitiveness?**

Governor Tarullo (2011) notes that the presence of scale economies at the largest financial institutions creates a trade-off between considerations of systemic risk and efficiency: "An additional concern would arise if some countries made the trade-off by limiting the size or configuration of their financial firms for systemic risk reasons at the cost of realizing genuine economies of scope or scale, while other countries did not. In this case, firms from the first group of countries might well be at a competitive disadvantage in the provision of certain cross-border activities."

Wheelock and Wilson (2010) explore this question by comparing the costs of the four largest financial institutions in the U. S. in 2009, which range in size from $1.244 trillion to $2.225 trillion, with the costs of a sufficient number of $1 trillion institutions whose total assets equal those of the four institutions. They find that the annual costs of these smaller institutions producing the same total assets as the four larger institutions would be about $20 billion higher per year, which is approximately 9 percent higher than the costs of the four largest institutions. The $1 trillion size of their smaller institutions still falls into the region considered too big to fail by many observers.

Taking $100 billion in total assets as the dividing line between too-big-to-fail banks and all others, Hughes and Mester (2011) compare the costs of the 17 institutions whose consolidated assets exceed $100 billion with a sufficient number of institutions scaled back in size to $100 billion so that they produce the same mix of financial products and their combined total assets equal those of the 17 largest ones. The estimated risk-return-driven cost function is used to predict the costs of these scaled-back $100 billion institutions. As a percent of consolidated assets, their total predicted costs are 10.9 percent higher than the estimated costs of the 17 institutions. Hence, the cost advantage of the 17 largest banks is substantial.

If larger banks are reduced in scale, what mix of products and services will they produce? If they produce the same mix as that of the largest banks observed today, their costs will be considerably higher and potentially uncompetitive in international markets. In a general equilibrium, they are unlikely to produce them if banks in other countries are able to achieve larger scale and proportionately lower costs.

**Conclusions**

Why is the evidence of scale economies limited and mixed – are they illusive or elusive? Investigations of the response of cost to an increase with scale that separate the effect on cost of scale-related technological advantages from the effect on cost of scale-related risk-taking uncover evidence of the elusive scale economies. Additional costs associated with increased risk-taking tend to obscure reduced costs associated with better diversification and other sources of scale economies.

If such economies exist, is it possible that they can be achieved by smaller institutions than the largest we observe today? When the effects of endogenous risk-taking are peeled away, the evidence shows that financial institutions of all sizes experience large scale economies, but the largest financial institutions obtain the largest scale economies.

Is the size of our largest financial institutions the result of technological cost advantages that improve the efficiency of their capital allocation and liquidity and enhance their international competitiveness? Or is their large size the result, not of technological cost advantages, but of safety-net subsidies that confer too-big-to-fail cost advantages and foster moral hazard in investment decisions? The estimated scale economies of the largest financial institutions are robust to the substitution of the interest rates paid on borrowed funds by smaller institutions which are not deemed too big to fail. Moreover, the deletion of the too-big-to-fail institutions from the estimation of the cost function and the prediction of their scale economies out of sample, from the smaller sample not subject to the too-big-to-fail doctrine, yield essentially the same large measured scale economies as the estimation from the full sample. These tests suggest that technological factors account for the scale economies of the largest banks.

The evidence of the role of technological factors in generating the largest scale economies at the largest financial institutions suggests that the trade-off between considerations of systemic risk and efficiency is genuine: the efficiency gains from large scale are not the result of cost advantages due to explicit and implicit safety-net subsidies. Consequently, proposals to restrict the size of the largest institutions must account for the implications of such restrictions on the international competitiveness of these institutions.

**Bibliography**


Berger, A.N. and Mester, L.J., 1997, "Inside the black box: what explains differences in the efficiencies of financial institutions," *Journal of Banking and Finance*, Vol. 21, pp. 895-947.

Bossone, B. and Lee, J.-K., 2004, "In finance, size matters: the 'systemic scale economies' hypothesis," IMF Staff Papers, 51:1.

Brewer, E. and Jagtiani, J., 2009, "How much did banks pay to become too-big-to-fail and to become systemically important?" Federal Reserve Bank of Philadelphia Working Paper No. 09-34.

DeYoung, R. E., Hughes, J. P., and Moon, C.-G., 2001, "Efficient risk-taking and regulatory covenant enforcement in a deregulated banking industry," *Journal of Economics and Business*, Vol. 53, pp. 255–82.

Feng, G. and Serletis, A., 2010, "Efficiency, technical change, and returns to scale in large US banks: Panel data evidence from an output distance function satisfying theoretical regularity," *Journal of Banking and Finance*, Vol. 34, pp. 127-138.

Financial Oversight Council, 2011, *Study of the Effects of Size and Complexity of Financial Institutions on Capital Market Efficiency and Economic Growth.*

Greenspan, Alan, 2010, "The crisis," *Brookings Papers on Economic Activity,* Spring, pp. 201-246.

Grossman, R.S., 1992, "Deposit insurance, regulations, and moral hazard in the thrift industry: evidence from the 1930's," *American Economic Review*, Vol. 82, pp. 800-821.

Hughes, J. P., 1999, "Incorporating risk into the analysis of production," Presidential Address, Atlantic Economic Society, *Atlantic Economic Journal*, Vol. 27, pp. 1-23.

Hughes, J.P., Lang, W., Mester, L.J. and Moon C.-G., 1996, "Efficient banking under interstate branching," *Journal of Money, Credit, and Banking*, Vol. 28, pp. 1045-1071.

Hughes, J.P., Lang, W., Mester, L.J. and Moon C.-G., 2000, "Recovering risky technologies using the almost ideal demand system: an application to U.S. banking," *Journal of Financial Services Research*, Vol. 18, pp. 5-27.

Hughes, J.P., Lang, W., Moon C.-G. and Pagano, M., 1997, "Measuring the efficiency of capital allocation in commercial banking," Working Paper 98-2, Federal Reserve Bank of Philadelphia (revised as Working Paper 2004-1, Rutgers University Economics Department).

Hughes, J.P. and Mester, L.J., 1998, "Bank capitalization and cost: evidence of scale economies in risk management and signaling," *Review of Economics and Statistics*, Vol. 80, pp. 314-325.

Hughes, J.P. and Mester, L.J., 2010, "Efficiency in banking: theory, practice, and evidence," Chapter 19 in *The Oxford Handbook of Banking*, edited by A.N. Berger, P. Molyneux, and J. Wilson, Oxford University Press.

Hughes, J.P. and Mester, L.J., 2011, "Who said large banks don't experience scale economies? Evidence from a risk-return-driven cost function," Wharton Financial Institutions Working Paper 11-47.

Hughes, J.P., Mester, L.J. and Moon C.-G., 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*, Vol. 25, pp. 2169-2208.

Keeley, M. C., 1990, "Deposit insurance, risk, and market power in banking," *American Economic Review*, Vol. 80, pp. 1183-1200.

Marcus, A.J., 1984, "Deregulation and bank financial policy," *Journal of Banking and Finance*, Vol. 8, pp. 557-565.

Tarullo, D., 2011, "Industrial organization and systemic risk: an agenda for further research," Conference on the Regulation of Systemic Risk, Federal Reserve Board, Washington, DC.

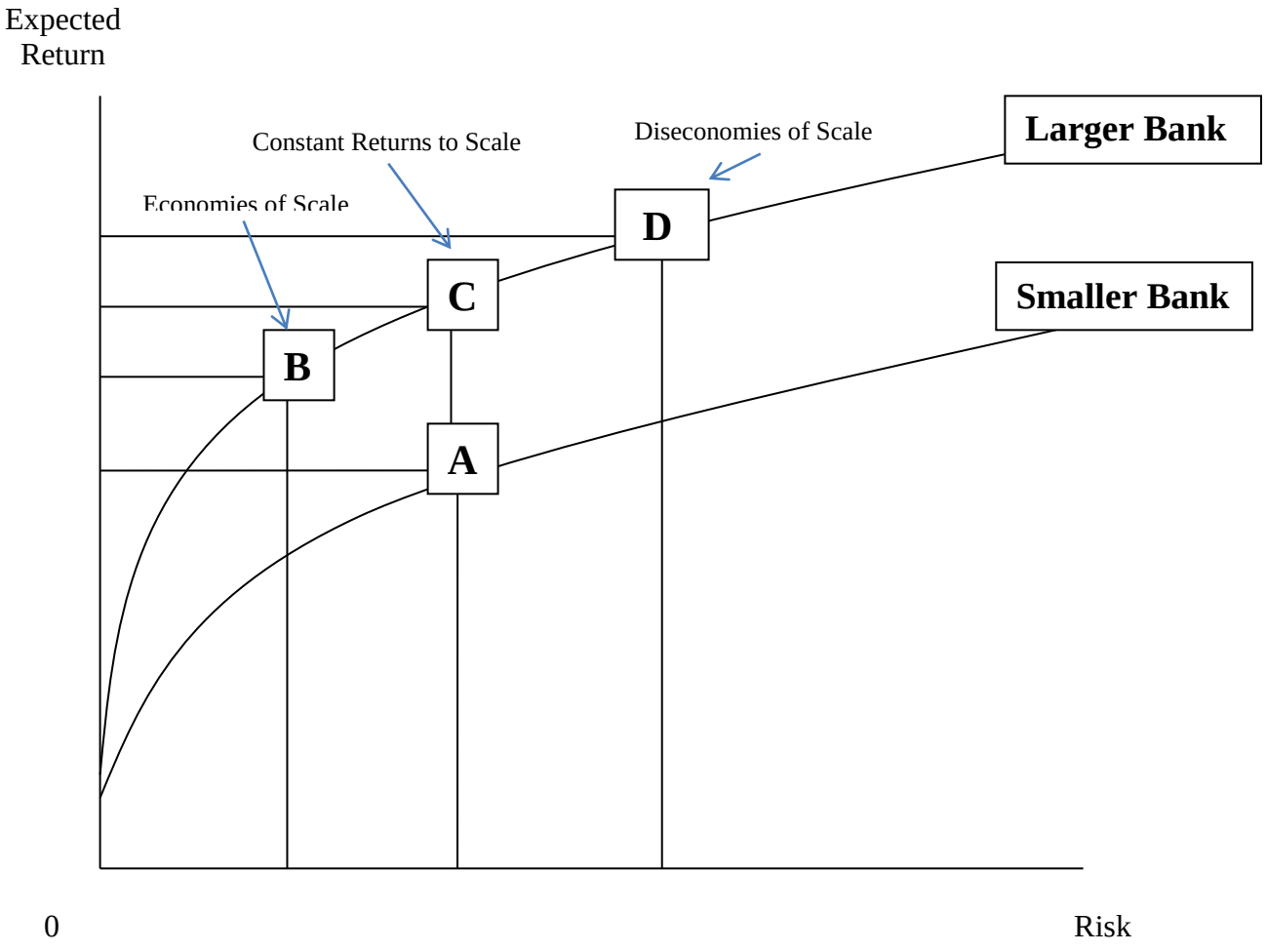
Tufano, P., 1996, "Who manages risk? An empirical examination of risk management practices in the gold mining industry," *Journal of Finance*, Vol. 50, pp. 1097-1137.

Wheelock, D. C. and P. W. Wilson., 2009 (revised 2010), "Do large banks have lower costs? New estimates of returns to scale for U. S. banks," Federal Reserve Bank of St. Louis, Working Paper 2009-054C.

## Figure 1

The lower frontier reflects the trade-off of a smaller, less diversified scale of operations. Consider a smaller institution whose mix of financial productions and services, asset quality, and funding strategies implies the trade-off at point A. Now consider a proportionally larger institution whose scale-related better diversification gives it the improved risk-return frontier. As the mix of financial products and services are increased, does cost increase less than proportionately, proportionately, or more than proportionately? If the larger institution takes no additional risks compared to the smaller institution at point A, the better diversification reduces its risk and improves its expected return, which is reflected by point B. Consequently, *cost at point B increases less than proportionately compared to point A, which implies economies of scale.* These scale economies reflect the better diversification of the underlying larger scale of operations.

Suppose, instead, that the larger institution takes advantage of the reduced marginal cost of risk management and takes additional risk for a higher expected return – say point C. At C, the risk of the larger institution equals that of the smaller one, but its expected return is higher. To the extent that the additional risk-taking results in higher costs than at point B, cost at point C may increase proportionately or even more than proportionately compared to point A, which, if the additional risk-taking is not taken into account, would obscure the inherent scale economies and give the appearance of constant returns to scale or even diseconomies of scale. At D, the larger institution takes more risk than the smaller institution so that the cost at point D is likely to increase more than proportionately compared to point A, which again gives the misleading appearance of diseconomies of scale. (These points were first made by Hughes (1999) and Hughes, Mester, and Moon (2001) and illustrated by this figure in Hughes and Mester (2011)).




Source: The figure and a similar discussion in the heading are found in Hughes and Mester (2011).

**Table 1**
**Estimated Mean Scale Economies for 1994**
**Reported as a Cost Elasticity**

Hughes, Mester, and Moon (2001) report these results in terms of the measure of scale economies which is the inverse of the cost elasticity. This table reports the cost elasticity instead. The cost elasticity gives the proportional response of cost to a proportional increase in all outputs. If cost increases less than proportionally, production is characterized by economies of scale: the value of cost elasticity in this case is less than one, and the value of scale economies, greater than one.

The data are taken from the Y9-C Call Reports filed quarterly with regulators. The sample included 441 top-tier U.S. bank holding companies in 1994.

The cost functions that generate the results in columns (1) and (2) are based on cost minimization. In column (1) and (3) cost sums all expenses except the cost of equity capital. The cost function in (2) includes the cost of equity. The function in (1) omits any role for equity capital while the function in (2) correctly includes equity as an argument and derives the shadow cost of equity. The risk-return-driven cost function in (3) accounts for endogenous risk-taking.

| | | **(1)** | **(2)** | **(3)** |
|---|---|---|---|---|
| | | **Mis-specified Cost Function** | **Economic Cost Function** | **Risk-Return-Driven Cost Function** |
| | | **Omits Level of Equity** | **Includes Shadow Cost of Equity** | **Conditioned on Optimal Equity** |
| **Total Assets** | **n** | **Mean** | **Mean** | **Mean** |
| **Full sample** | 441 | 0.9885** | 1.0187* | 0.8737*** |
| **< $0.3 billion** | 109 | 0.9882 | 1.0230* | 0.8949*** |
| **$0.3 billion – $2 billion** | 215 | 0.9878** | 1.0167* | 0.8883*** |
| **$2 billion – $10 billion** | 67 | 0.9887* | 1.0183* | 0.8537*** |
| **$10 billion – $50 billion** | 35 | 0.9924 | 1.0223* | 0.8017*** |
| **> $50 billion** | 12 | 0.9922 | 1.0130 | 0.7998*** |

All estimates of mean cost elasticities are significantly different from 0 at the 1 percent level.
* Significantly different from 1 at the 10 percent level
** Significantly different from 1 at the 5 percent level
*** Significantly different from 1 at the 1 percent level

**Table 2**
**Estimated Mean Scale Economies for 1990 Compared to 1994**
**Reported as a Cost Elasticity**

Hughes, Lang, Mester, and Moon (2000) report the results in columns (1) and (2) for the 286 U.S. commercial banks in 1990 that exceed $1 billion in total assets. Column (3) reports the cost elasticities for U.S. bank holding companies in 1994 derived by Hughes, Mester, and Moon (2001) for the size groups (in 1990 dollars) defined in study using 1990 data. The bank data are obtained from the Consolidated Reports of Condition and Income filed quarterly with regulators. The data from Hughes, Mester, and Moon (2001) are taken from the Y9-C Call Reports filed quarterly with regulators. The full sample included 441 top-tier U.S. bank holding companies in 1994, but there are 151 institutions in the size range defined by the 1990 data.

Both studies report the measure of scale economies which is the inverse of the cost elasticity. This table reports the cost elasticity instead. The cost elasticity gives the proportional response of cost to a proportional increase in all outputs. If cost increases less than proportionally, production is characterized by economies of scale: the value of cost elasticity in this case is less than one, and the value of scale economies, greater than one.

The cost function that generates the results in column (1) is based on cost minimization. The measure of costs in columns (1), (2), and (3) sums all costs except the cost of equity capital. The risk-return-driven cost function in columns (2) and (3) accounts for endogenous risk-taking.

| | | **(1)** | **(2)** | | **(3)** |
|---|---|---|---|---|---|
| | | **1990** | **1990** | **1994** | **1994** |
| | | **Mis-specified Cost Function** | **Risk-Return-Driven Cost Function** | **BHCs in the size ranges defined by 1990 study** | **Risk-Return-Driven Cost Function** |
| | | **Omits Level of Equity** | **Conditioned on Optimal Equity** | | **Conditioned on Optimal Equity** |
| **Total Assets in 1990 dollars** | **n** | **Mean** | **Mean** | **n** | **Mean** |
| **Full sample** | 286 | 0.967*** | 0.877*** | 151 | 0.856*** |
| **$1.00 billion – $1.75 billion** | 72 | 0.979*** | 0.908*** | 40 | 0.895*** |
| **$1.75 billion – $3.00 billion** | 71 | 0.972*** | 0.887*** | 27 | 0.883*** |
| **$3.00 billion – $6.40 billion** | 72 | 0.966*** | 0.873*** | 27 | 0.842*** |
| **$6.40 billion – $70.00 billion** | 71 | 0.952*** | 0.828*** | 57 | 0.824*** |

All estimates of mean cost elasticities are significantly different from 0 at the 1 percent level.
* Significantly different from 1 at the 10 percent level
** Significantly different from 1 at the 5 percent level
*** Significantly different from 1 at the 1 percent level