WORKING PAPERS
RESEARCH DEPARTMENT

WORKING PAPER NO. 11-27

WHO SAID LARGE BANKS DON'T EXPERIENCE SCALE ECONOMIES? EVIDENCE FROM A RISK-RETURN-DRIVEN COST FUNCTION

Joseph P. Hughes
Department of Economics, Rutgers University

Loretta J. Mester
Research Department, Federal Reserve Bank of Philadelphia
and
Finance Department, The Wharton School, University of Pennsylvania

July 2011

RESEARCH DEPARTMENT, FEDERAL RESERVE BANK OF PHILADELPHIA

Ten Independence Mall, Philadelphia, PA 19106-1574 • www.philadelphiafed.org/research-and-data/

WHO SAID LARGE BANKS DON'T EXPERIENCE SCALE ECONOMIES? EVIDENCE FROM A RISK-RETURN-DRIVEN COST FUNCTION

Joseph P. Hughes
Department of Economics, Rutgers University

Loretta J. Mester
Research Department, Federal Reserve Bank of Philadelphia
and
Finance Department, The Wharton School, University of Pennsylvania

July 2011

Abstract

Earlier studies found little evidence of scale economies at large banks; later studies using data from the 1990s uncovered such evidence, providing a rationale for very large banks seen worldwide. Using more recent data, we estimate scale economies using two production models. The standard risk-neutral model finds little evidence of scale economies. The model using more general risk preferences and endogenous risk-taking finds large scale economies. We show that these economies are not driven by too-big-to-fail considerations. We evaluate the cost implications of breaking up the largest banks into banks of smaller size.

*The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of Philadelphia or of the Federal Reserve System.

We thank Marc Rodriguez and Arland Kane of the Board of Governors of the Federal Reserve System for their help with the data, Sally Burke for editorial assistance, and Christian Laux, conference participants at the International Banking, Economics, and Finance Association session at the Allied Social Sciences Association meetings in January 2011, and conference participants at the Financial Intermediation Research Society Annual Meeting in June 2011 for very helpful comments.

Correspondence to Hughes at Department of Economics, Rutgers University, New Brunswick, NJ 08901-1248; phone: (917) 721-0910; email: jphughes@rci.rutgers.edu. To Mester at Research Department, Federal Reserve Bank of Philadelphia, Ten Independence Mall, Philadelphia, PA 19106-1574; phone: (215) 574-3807; email: Loretta.Mester@phil.frb.org.

JEL Codes: D20, D21, G21, L23.
Key Words: banking, production, risk, scale economies, too big to fail.

For years the Federal Reserve was concerned about the ever-growing size of our largest financial institutions. Federal Reserve research had been unable to find economies of scale in banking beyond a modest size.

Alan Greenspan
"The Crisis" (*Brookings Papers on Economic Activity*, Spring 2010, p. 231)

I. Introduction

The financial crisis that began in 2007 has focused attention on large financial institutions and the role the too-big-to-fail doctrine has played in driving their size. Financial reform has focused on limiting the costs that systemically important financial institutions (SIFIs) impose on the economy. However, the potential efficiency benefits of operating at a large scale have been largely neglected in policy discussions and recent research. Textbooks explain that banks should enjoy scale economies as they grow larger because the credit risk of their loans and financial services, as well as the liquidity risk of their deposits, becomes better diversified, which reduces the relative cost of managing these risks and allows banks to conserve equity capital as well as reserves and liquid assets. In addition, textbooks point to the spreading of overhead costs, especially those associated with information technology, as another source of scale economies. But the recent financial crisis has led many to question whether such efficiencies exist or whether scale has been driven primarily by institutions seeking to exploit the cost advantages of being too big to fail.

Older empirical studies that used data from the 1980s did not find scale economies in banking except at very small banks. But more recent studies that used data from the 1990s and 2000s and more modern methods for modeling bank technology that incorporate managerial preferences for risk and endogenize bank risk-taking find significant scale economies at banks of all sizes included in the sample.[1] These studies include Hughes, Lang, Mester, and Moon (1996, 2000), Berger and Mester (1997), Hughes and Mester (1998), Hughes, Mester, and Moon (2001), Bossone and Lee (2004), Wheelock and Wilson (2009), and Feng and Serletis (2010). Hughes and Mester (2010) discuss some of these modern methods

[1] See Mester (2010) for discussion.

of modeling bank technology and the evidence of scale economies obtained from them. Part of the difference in results between the older studies and more recent ones appears to reflect improvements in the methods researchers used for measuring scale economies and part reflects a change in banking technology, such as the use of information technologies, and environmental factors, such as geographic deregulation, which have led to a larger efficient scale of banking production.

This investigation uses the modeling techniques developed by Hughes, Lang, Mester, and Moon (1996, 2000), and Hughes, Mester, and Moon (2001). These earlier papers used 1994 data on top-tier bank holding companies in the U.S., while here we use 2007 data.[2] During the thirteen years that separate these two data sets, advances in information technology and further implementation of this technology in banking as well as greater diversification from geographic consolidation might be expected to increase economies of scale in banking. And indeed, consistent with the textbook prediction and with consolidation in the banking industry, we find large scale economies at small banks and even larger scale economies at large banks. The finding of significant scale economies even at banks that are not at a size usually considered too big to fail suggests that government policy is not the only source of size-related cost economies. We present evidence below that too-big-to-fail considerations are not the source of the scale economies we find. In addition, we provide estimates of the cost impact of breaking up banks into smaller institutions, as has been proposed by some. In performing this exercise we take into account not only the size of the banks but also the potential longer-run impact that accounts for the fact that smaller banks focus on different product offerings than larger banks. Our results suggest that while reducing the size of banks would raise the costs of production holding output mix constant (the scale effect), once these banks adjusted their product mix, there would be cost savings. Whether this is socially beneficial, however, depends on whether the product mix offered by the largest banks was beneficial – a question that is beyond the scope of the current paper.

[2] The BHCs in our data set range in size from $72 million to $2.19 trillion in total assets. We performed additional tests that show that our results are robust to estimating the model excluding the largest banks in the sample, estimating the model excluding smaller banks in the sample, and estimating the model excluding extreme values for the output shares.

The paper proceeds as follows. Sections II-IV discuss the theoretical model that incorporates bank managers' risk preferences and endogenous choice of risk. Those mainly interested in the empirical results can skip to Sections V-VI, which discuss the empirical model specifications. Section VII discusses our data set. Sections VIII-X give our empirical results, and Section XI concludes.

II. Modeling Banking Cost

According to the standard textbook, a cost function uses input prices to translate the production function into the minimum cost of producing output. The textbook usually illustrates the cost function in terms of an expansion path graphed on an isoquant map. The expansion path is the locus of points where the marginal rate of substitution equals the input price ratio. The older literature on modeling bank cost functions often applied these concepts in a very straightforward way to bank production. It considered how to specify outputs and inputs in terms of bank assets, financial services, and liabilities. After calculating input prices, it derived a cost function for econometric estimation, applied it to bank data, and computed scale economies from the fitted function. As noted above, the results usually offered no evidence of scale economies at large banks.

Hughes, Lang, Mester, and Moon (1996, 2000), and Hughes, Mester, and Moon (2001) argue that the standard specification of the cost function fails to capture an essential ingredient in bank production – *risk*. Bank managers' risk preferences are typically not modeled in standard cost function analysis, yet managers face a risk-expected-return trade-off determined by the investment strategy they choose and the economic environment in which they operate. Thus, a bank's cost depends on its risk exposure, which contains an exogenous component reflecting the economic environment and an endogenous component reflecting the managers' choice of risk exposure.

The standard textbook explains that banks might enjoy scale economies derived from the diversification of risk obtained from a larger portfolio of loans and a larger base of deposits. These diversification benefits allow larger banks to manage risk with relatively fewer resources. In other words, a larger scale of operations improves a bank's risk-return trade-off. Figure 1 shows the smaller bank's investment strategies on the risk-return frontier labeled *I* and the larger bank's strategies on frontier *II*. Suppose in Figure 1 point *A* represents production of a smaller, less diversified output, say, some quantity of loans with a particular probability distribution of default that reflects the contractual interest rate charged and the resources allocated to risk assessment and monitoring. Point *B* represents a larger quantity of loans with the same contractual interest rate but better diversification and, hence, an improved probability distribution of default and lower overall risk. The better diversification allows the costs of risk management to increase less than proportionately with the loan volume while maintaining an improved probability distribution of default. Thus, the response of cost to the increase in output from point A to point B reflects scale economies and the expected return at *B* exceeds that at *A*.

Suppose, instead, the larger, better diversified portfolio of loans is produced with the investment strategy at point *C*. The strategy at *C* preserves the risk exposure of *A*, and the better diversification improves the expected return. The bank at *C* may charge a higher contractual interest rate, which would tend to increase risk, but the better diversification offsets the additional risk. The cost of managing the larger loan portfolio at the same risk as *A* may still increase less than proportionately, but the increase will be greater than that occasioned by *B*. Thus, the change in cost from *A* to *C* may still show scale economies, though smaller than from *A* to *B*.

On the other hand, suppose the bank responds to the better diversification of the larger output by adopting a more risky investment strategy for an enhanced expected return. It charges an even higher contractual interest rate on loans than at point *C*. Better diversification does not offset the increased cost occasioned by the additional default risk. Point *D* in Figure 1 designates this strategy. The increased inherent default risk due to the higher contractual interest rate results in costs of risk management that increase more than proportionately with the loan volume (from *A* to *D*), and production appears to exhibit the counter-intuitive *scale diseconomies* found by empirical studies of banking cost that fail to account for endogenous risk-taking.

While the investment strategies at *B*, *C*, and *D* entail producing the *same quantity of loans*, the expected return and its associated cost and risk of producing the loans differ across the three strategies.

Figure 2 illustrates this point. It characterizes the production technology for a quantity of loans represented by the isoquant shown in the figure. The mix of debt and equity used to fund the loans is ignored. Instead, the diagram shows the quantity of physical capital and labor used in the process of credit evaluation and loan monitoring. (As the argument that follows illustrates, this isoquant is not well defined in traditional terms.) Point *C* shows the least costly way to produce the particular quantity of loans with the risk exposure associated with the investment strategy *C* in Figure 1. If a bank adopted the less risky strategy, *B*, it might use less labor in credit evaluation and monitoring: point *B* in Figure 2, a less costly method of producing the *same quantity of loans*. Thus, the isoquant for this quantity of loans that passes through point *C* captures one investment strategy only. If the isoquant included a characterization of the risk exposure, there would be another isoquant passing through point *B* for the same quantity of loans produced with the lower risk strategy. On the other hand, if a bank adopted the more risky strategy, *D*, it would use more labor, the corresponding point *D* in Figure 2, a more costly method than *C*. Thus, the cost of producing this particular quantity of loans depends on a bank's choice of risk exposure and its expected return. As in Hughes, Lang, Mester, and Moon (1996, 2000), and Hughes, Mester, and Moon (2001), we refer to this risk-return-driven cost as the managerial most-preferred cost function, since it reflects managers' preferences over investment strategies that reflect the risk-expected return trade-off.

As explained in Hughes, Mester, and Moon (2001), failing to account for endogenous risk-taking when estimating a production model can produce misleading estimates of scale economies and cost elasticities. If production is observed at the points *A* and *B*, a naïve calculation of the cost elasticity from the difference in cost measured at these two points would appear to yield evidence of scale economies. If production is observed at points *A* and *D*, a naïve calculation from their difference in cost would appear to give evidence of scale diseconomies. Thus, the specification of the cost function to be estimated must account for endogenous risk-taking to detect the scale economies associated with better diversification.

III. Modeling Managers' Preferences for Return and Risk

Figures 1 and 2 illustrate that the cost of producing the larger, better diversified output depends on managers' choice of investment strategy in response to the better risk-expected-return trade-off. Thus, cost is not independent of managers' risk preferences. Why might risk influence banks' production choices?

Modern banking theory emphasizes that bank managers face dichotomous investment strategies for maximizing value: one, higher risk; the other, lower risk (Marcus, 1984). The higher risk strategy, characterized in part by a lower capital ratio and lower asset quality, exploits mispriced deposit insurance, too-big-to-fail policies, and other benefits of the governmental safety net. Of course, this strategy also increases the risk of financial distress – possibly involving regulatory intervention in the operations of the bank, liquidity crises, and even insolvency and loss of the bank's charter. Such a risky strategy enhances a bank's value when its investment opportunities are not particularly valuable: the expected gains from exploiting safety-net subsidies outweigh the potential losses entailed in episodes of financial distress. On the other hand, if a bank enjoys valuable investment opportunities, these market advantages increase its expected costs of financial distress. When the expected losses involved in financial distress exceed the expected gains from exploiting the safety net, banks enhance their value by pursuing a lower-risk strategy involving a higher capital ratio and higher asset quality.[3] Both of these investment strategies maximize firm value. Hence, risk-neutral managers would pursue them. They manage risk when doing so maximizes value (Tufano, 1996).

These value-maximizing, dichotomous investment strategies highlight the importance of accounting for endogenous risk-taking in estimating production costs in banking. Modeling managers' risk preferences forms the foundation for building a model of bank production and cost.

We turn first to some notational matters. We represent bank technology by the transformation function, $T(\boldsymbol{y}, n, p, \boldsymbol{x}, k) \leq 0$, where $\boldsymbol{y}$ denotes information-intensive loans and financial services; k, equity

---

[3] For empirical evidence of these dichotomous strategies, see Keeley (1990) and Hughes, Lang, Moon, and Pagano (1997, 2004).

capital; $\boldsymbol{x}_d$, demandable debt and other types of debt; $\boldsymbol{x}_b$, labor and physical capital; and $\boldsymbol{x} = (\boldsymbol{x}_b, \boldsymbol{x}_d)$. The price of the *i*-th type of input is designated by w_i so that the economic cost of producing the output vector $\boldsymbol{y}$ is given by $\boldsymbol{w_b x_b} + \boldsymbol{w_d x_d} + w_k k$. If the cost of equity capital is omitted, $\boldsymbol{w_b x_b} + \boldsymbol{w_d x_d}$ gives the *cash-flow cost* (C_{CF}). We characterize asset quality by two types of proxies: *ex ante* measures are given by the vector of average contractual interest rates on assets such as securities and loans, $\boldsymbol{p}$, which, given the risk-free interest rate, r, captures an average risk premium, and an *ex post* measure, the amount of nonperforming loans, n.

Rather than express managers' preferences in terms of how they rank expected return and return risk, the first two moments of the subjective distribution of returns, we ask how managers rank production plans. Production plans are more basic: to rank production plans, managers must translate plans into subjective, conditional probability distributions of profit. Managers' beliefs about the probability distribution of states of the world, s_t, and about how the interaction of production plans with states yields a realization of after-tax profit, $\pi = g(\boldsymbol{y}, n, \boldsymbol{p}, r, \boldsymbol{x}, k, s)$, imply a subjective distribution of profit that is conditional on the production plan: $f(\pi; \boldsymbol{y}, n, \boldsymbol{p}, r, \boldsymbol{x}, k)$. Under certain restrictive conditions, this distribution can be represented by its first two moments, $E(\pi; \boldsymbol{y}, n, \boldsymbol{p}, r, \boldsymbol{x}, k)$ and $S(\pi; \boldsymbol{y}, n, \boldsymbol{p}, r, \boldsymbol{x}, k)$. Rather than define a utility function over these two moments, we define it over profit and the production plan, $U(\pi; \boldsymbol{y}, n, \boldsymbol{p}, r, \boldsymbol{x}, k)$, which is equivalent to defining it over the conditional probability distributions $f(\cdot)$. This generalized managerial utility function subsumes the case of profit maximization where only the first moment of the conditional distribution of profit influences utility; however, it also explains cases where higher moments influence utility so that managers can trade profit to achieve other objectives involving risk.

IV. Modeling Cost When Risk Is Endogenous

The cost of producing a particular output vector $\boldsymbol{y}$ – financial assets and services – depends on the employment of inputs $\boldsymbol{x}$ and k – labor, physical capital, debt, and equity. How managers choose to produce any particular output vector can be modeled as a utility maximization problem. Hence, the choice from the production strategies highlighted in Figures 1 and 2, points *B*, *C*, and *D*, solves the utility maximization problem.

Since the utility function ranks production plans – output and input vectors and the resulting profit – banks maximize utility conditional on the output vector by solving for the utility-maximizing profit and the constituent vector of inputs required to produce it. Let m designate noninterest income while $\boldsymbol{p} \cdot \boldsymbol{y}$ represents interest income. Total revenue is given by $\boldsymbol{p} \cdot \boldsymbol{y} + m$. Letting π designate after-tax profit and t, the tax rate on profit, and $p_\pi = 1/(1-t)$, the price of a dollar of after-tax profit in terms of before-tax dollars, the before-tax accounting or cash-flow profit is defined as $p_\pi \pi = \boldsymbol{p} \cdot \boldsymbol{y} + m - \boldsymbol{w_b} \cdot \boldsymbol{x_b} - \boldsymbol{w_d} \cdot \boldsymbol{x_d}$.

The utility-maximization problem is given by:

$$\max_{\pi,\, \boldsymbol{x}} U(\pi, \boldsymbol{x}; \boldsymbol{y}, n, \boldsymbol{p}, r, k) \tag{1a}$$

$$\text{s.t.}\quad p_\pi \pi = \boldsymbol{p} \cdot \boldsymbol{y} + m - \boldsymbol{w_b} \cdot \boldsymbol{x_b} - \boldsymbol{w_d} \cdot \boldsymbol{x_d} \tag{1b}$$

$$T(\boldsymbol{y}, n, \boldsymbol{p}, r, \boldsymbol{x}, k) \leq 0. \tag{1c}$$

The solution gives the *managers' most preferred profit function*, $\pi^* = \pi_{MP}(\boldsymbol{y}, n, \boldsymbol{v}, k)$, and the *managers' most preferred input demand functions*, $\boldsymbol{x}^* = \boldsymbol{x}_{MP}(\boldsymbol{y}, n, \boldsymbol{v}, k)$, where $\boldsymbol{v} = (\boldsymbol{w}, \boldsymbol{p}, r, m, p_\pi)$. The *managers' most preferred cost function* follows trivially from the profit function:

$$C_{MP}(\boldsymbol{y}, n, \boldsymbol{v}, k) = \boldsymbol{p} \cdot \boldsymbol{y} + m - p_\pi \pi_{MP}(\boldsymbol{y}, n, \boldsymbol{v}, k). \tag{2}$$

We claimed above that this utility-maximization problem has sufficient structure to identify and control for the choice of production plan from points *B*, *C*, and *D* of Figures 1 and 2 – plans that produce the same output, $\boldsymbol{y}$, but differ in their risk exposure and resources allocated to managing risk. How then does the solution, the most preferred profit and cost functions and the most preferred input demand functions, depend on the risk exposure?

First, note that revenue, $\boldsymbol{p} \cdot \boldsymbol{y} + m$, drives the solution. In addition, the output prices, $\boldsymbol{p}$, which are contractual returns on assets such as loans and securities, control for the *ex ante* risk premium of each of those assets when they are compared to the risk-free interest rate r. The quantity of nonperforming assets, n, captures *ex post* or realized default risk. The quantity of equity capital, k, controls for a key component of capital structure that underlies expected return and return risk. Moreover, since the cost of equity and loan losses are excluded from the calculation of cash-flow cost and profit, the quantities of equity and nonperforming loans control for these omitted expenses. These controls as well as the tax rate on earnings embodied in the price of a before-tax dollar, p_π, in terms of after-tax dollars constitute a rich characterization of investment strategies that shape cost.

These variables that characterize and control for the investment strategy permit the calculation of risk-adjusted scale economies from the estimated cost function – a calculation that accounts for the bank's choice of risk exposure. In Figures 1 and 2 the problem of identifying the points B, C, and D for the purpose of computing scale economies is resolved by these control variables. Note that to the extent that larger banks and smaller banks choose a different product mix with different risk characteristics – e.g., larger banks produce more off-balance-sheet activities than smaller banks – by controlling for risk preferences, this cost model allows us to include banks of all sizes in our estimation.

V. Using the Almost Ideal Demand System to Estimate the Most Preferred Cost Function and Scale Economies

To estimate the utility-maximizing profit and input demand functions that solve the problem (1a), (1b), and (1c), we follow Hughes, Lang, Mester, and Moon (1996, 2000), and Hughes, Mester, and Moon (2001) and adapt the Almost Ideal Demand System of consumer theory, which was proposed by Deaton and Muellbauer (1980), to represent managerial preferences. Just as the estimation of this system using consumers' budget data recovers consumers' preferences for goods and services, its application to banks' data on production and cost recovers managers' rankings of production plans or, equivalently, their ranking of subjective probability distributions of profit conditional on the production plan.

The profit equation and input demands are expressed as expenditure shares of total revenue:

(3a) $$\frac{p_\pi \pi}{\mathbf{p}\cdot\mathbf{y}+m} = \frac{\partial \ln \boldsymbol{P}}{\partial \ln p_\pi} + \mu[ln(\mathbf{p}\cdot\mathbf{y}+m) - \ln \boldsymbol{P}]$$

(3b) $$\frac{w_i x_i}{\mathbf{p}\cdot\mathbf{y}+m} = \frac{\partial \ln \boldsymbol{P}}{\partial \ln w_i} + \nu_i[ln(\mathbf{p}\cdot\mathbf{y}+m) - \ln \boldsymbol{P}]\ \forall i$$

where $ln\ \mathbf{P} = A_0 + \Sigma_i A_i\ ln\ y_i + (½)\Sigma_i \Sigma_j S_{ij}\ ln\ y_i\ ln\ y_j + \Sigma_i B_i\ ln\ w_i + (½)\Sigma_i \Sigma_j G_{ij}\ ln\ w_i\ ln\ w_j$

$+ \Sigma_i \Sigma_j D_i\ ln\ y_i\ ln\ w_j + (½)\Sigma_i \Sigma_j R_{ij}\ ln\ z_i\ ln\ z_j + \Sigma_i \Sigma_j H_{ij}\ ln\ z_i\ ln\ y_j + \Sigma_i \Sigma_j T_{ij}\ ln\ z_i\ ln\ w_j$;

and $\mathbf{z} = (k, n, \boldsymbol{p}, p_\pi)$.

The input shares and profit share sum to one.

Equity capital enters the specification of the profit and input demand equations as a conditional argument. Hence, we include in the estimation a first-order condition defining the utility-maximizing value of equity capital:

(3c) $$\frac{\partial V(\cdot)}{\partial k} = \frac{\partial V(\cdot)}{\partial \ln k}\frac{\partial \ln k}{\partial k} = 0,$$

where the indirect utility function, $V(\cdot)$, of the maximization problem (1a)-(1c) is

(4) $$V(\cdot) = \frac{\ln(\mathbf{p}\cdot\mathbf{y}+m) - \ln \boldsymbol{P}}{\beta_0\left(\prod_i y_i^{\beta_i}\right)\left(\prod_j w_i^{\nu_i}\right)p_\pi{}^{\mu} k^{\kappa}}.$$

The appendix gives the details of empirical specification and estimation.

Managerial preferences represent their beliefs about the probabilities of future states of the world and how those states interact with production plans to generate realizations of profit, so managers' preferences change over time. Consequently, we use cross-sectional data and estimate the production system with nonlinear two-stage least squares, which is a generalized method of moments.

Hughes, Lang, Mester, and Moon (1996, 2000) show that the Almost Ideal profit share equation is identically equal to the translog profit (cost) function when the comparative-static restrictions implied by the assumption of profit maximization are imposed or when they are satisfied by the data. Thus, these restrictions provide a test for the consistency of the data with profit maximization (cost minimization).

Equation 2 shows how the managers' most preferred cost function is derived from the profit function. We compute the measure of scale economies, the inverse of the cost elasticity with respect to output, from this expression after substituting the optimal demand for equity capital into it:

$$\text{(5)} \quad \begin{aligned} \textit{most preferred cost economies} &= \frac{C_{MP}}{\sum_i y_i \left[\dfrac{\partial C_{MP}}{\partial y_i} + \dfrac{\partial C_{MP}}{\partial k} \dfrac{\partial k}{\partial y_i} \right]} \\ &= \frac{\mathbf{p} \cdot \mathbf{y} + m - p_\pi \pi}{\sum_i y_i \left[p_i - \dfrac{\partial p_\pi \pi}{\partial y_i} - \dfrac{\partial p_\pi \pi}{\partial k} \dfrac{\partial k}{\partial y_i} \right]} , \end{aligned}$$

A value greater than one implies scale economies, and a value less than one implies scale diseconomies.

VI. Minimum Cost Functions and Scale Economies

The standard minimum cost function is quite different from the most preferred cost function just discussed. The standard cost function can control for some aspects of risk, including the amount of nonperforming loans, *n*, which accounts for the influence of asset quality on cost. In addition, the important role of equity capital in banking production suggests that the minimum cost function should include either the required return (price) or quantity of equity capital. When the required return is not readily available (and it isn't, since most banks are not publicly traded), the minimum cost function can be conditioned on equity capital. In this case, the cost function excludes the cost of equity capital and, thus, is *cash-flow* cost. Note that this function fails to account for the revenue side of expected return and return risk that are found in the specification of the most preferred profit and cost functions. Thus, the standard cash-flow cost function is:

$$C_{CF}(\boldsymbol{y}, n, \boldsymbol{w_b}, \boldsymbol{w_d}, k) = \min_{\boldsymbol{x_b}, \boldsymbol{x_d}} (\boldsymbol{w_b} \cdot \boldsymbol{x_b} + \boldsymbol{w_d} \cdot \boldsymbol{x_d}) \text{ s.t. } T(\boldsymbol{y}, n, \boldsymbol{x}, k) \leq 0 \text{ and } k = k^0. \tag{6}$$

We estimate this cost function and its associated share equations with a translog specification:

$$\ln C_{CF} = \alpha_0 + \Sigma_i \alpha_i \ln g_i + (½)\Sigma_{ij} \alpha_{ij} \ln g_i \ln g_j \text{ and } \boldsymbol{g} = (\boldsymbol{y}, n, \boldsymbol{w}, k).$$

Scale economies based on this cash-flow cost function are:

$$\textit{cash-flow scale economies from the } C_{CF} \textit{ cost function} = \frac{1}{\sum_i \frac{\partial \ln C_{CF}}{\partial \ln y_i}}. \tag{7}$$

Some studies of banking technology neglect the critical role of equity capital by defining a minimum cash-flow cost function without conditioning it on the amount of equity capital:

$$C_{MS}(\boldsymbol{y}, n, \boldsymbol{w_b}, \boldsymbol{w_d}) = \min_{\boldsymbol{x_b}, \boldsymbol{x_d}} (\boldsymbol{w_b} \cdot \boldsymbol{x_b} + \boldsymbol{w_d} \cdot \boldsymbol{x_d}) \text{ s.t. } T(\boldsymbol{y}, n, \boldsymbol{x}) \leq 0. \tag{8}$$

To illustrate the bias introduced by such a cash-flow cost function, consider two banks identical in every respect except their capital structures. One bank uses less equity and more debt to finance the same quantity of assets. Thus, its cash-flow cost of producing the same output will be greater because it incurs the interest cost of the additional debt. Since cash-flow cost does not account for the cost savings of less equity, it appears to be a more costly method of producing the same output. Had the cash-flow cost function been properly conditioned on the amount of equity capital employed, the appearance of a less efficient production method would have been dispelled. Thus, the specification of cost in (8) is theoretically mis-specified so we label it with the *MS* subscript. For illustrative purposes, we estimate this cost function and its associated share equations with a translog specification:

$$ln\ C_{\mathbf{MS}} = \alpha_0 + \Sigma_i\, \alpha_i\, ln\ h_i + (½)\Sigma_i\, \Sigma_j\, \alpha_{ij}\, ln\ h_i\, ln\ h_j \text{ and } \boldsymbol{h} = (\boldsymbol{y}, n, \boldsymbol{w}).$$

Scale economies based on this cost function are given by:

$$(9) \qquad \textit{cash-flow scale economies from the } C_{MS} \textit{ cost function} = \frac{1}{\sum_i \frac{\partial \ln C_{MS}}{\partial \ln y_i}} .$$

In contrast to these cash-flow cost functions, consider an economic cost function that includes the cost of equity capital:

$$(10) \qquad C_{\mathbf{EC}}(\boldsymbol{y}, n, \boldsymbol{w_b}, \boldsymbol{w_d}, w_k) = \min_{\boldsymbol{x_b}, \boldsymbol{x_d}, k} (\boldsymbol{w_b} \cdot \boldsymbol{x_b} + \boldsymbol{w}_d \cdot \boldsymbol{x}_d + w_k k) \text{ s.t. } T(\boldsymbol{y}, n, \mathbf{x}, k) \le 0.$$

Since the economic cost function includes the cost of equity capital, it is conditioned on the required return (price) rather than the quantity of equity capital. When a bank is publicly traded, the required return, w_k, can be computed from an asset pricing model; however, most banks are not publicly traded. Instead, the cash-flow cost function in (6) is used to obtain a shadow price of equity capital from which the economic cost function and its associated scale economies can be computed.[4] The first-order condition for optimal equity capital gives its shadow price:

$$(11) \qquad w_k = -\frac{\partial C_{CF}}{\partial k} .$$

Then the economic cost function is:

$$(12) \qquad C_{EC}(y, n, \boldsymbol{w_b}, \boldsymbol{w_d}, w_k) = \min_k C_{CF}(y, n, \boldsymbol{w_b}, \boldsymbol{w_d}, k) + w_k k = \min_k C_{CF}(y, n, \boldsymbol{w_b}, \boldsymbol{w_d}, k) - \frac{\partial C_{CF}}{\partial k} k.$$

[4] Braeutigam and Daughety (1983) first suggested this technique, and Hughes, Mester, and Moon (2001) applied it to banking production and cost.

If we assume that the observed level of equity capital is cost-minimizing, then marginal cost computed from the cash-flow cost function equals marginal cost computed from the economic cost function:[5]

$$\frac{\partial C_{EC}}{\partial y_i} = \frac{\partial C_{CF}}{\partial y_i} \;\forall i. \tag{13}$$

Then, using (12) and (13), the degree of scale economies based on the economic cost function is given by:

$$\textit{economic cost scale economies from the } C_{EC} \textit{ cost function} = \frac{1}{\sum_i \frac{\partial \ln \mathrm{C}_{EC}}{\partial \ln y_i}} \tag{14}$$

$$= \frac{\mathrm{C}_{EC}}{\sum_i y_i \frac{\partial\, \mathrm{C}_{EC}}{\partial\, y_i}} = \frac{\mathrm{C}_{CF} - k\frac{\partial\, \mathrm{C}_{CF}}{\partial\, k}}{\sum_i y_i \frac{\partial\, \mathrm{C}_{CF}}{\partial\, y_i}} = \frac{1 - \frac{\partial \ln \mathrm{C}_{CF}}{\partial \ln k}}{\sum_i \frac{\partial \ln \mathrm{C}_{CF}}{\partial \ln y_i}}.$$

VII. The Data

Our data set includes 842 top-tier bank holding companies in the United States in 2007. A top-tier company is not owned by another company. The data are obtained from the Y-9 C Call Reports filed quarterly with bank regulators. We model the consolidated bank rather than its constituent banks and subsidiaries because investment decisions are generally made at the consolidated level. The summary statistics describing these banks are found in Tables 1-5.

In all the cost functions we specify the same five outputs: y_1, comprising cash, repos, federal funds sold, and interest-bearing deposits due from banks; y_2, securities, including U.S. Treasury and U.S. government agency securities as well as nongovernmental securities; y_3, loans; y_4, trading assets,

---

[5] Interpreting this proposition in terms of long-run (economic) cost and short-run variable (cash-flow) cost, it illustrates the familiar result that long-run and short-run marginal costs are equal when the value of the "fixed" input that gives rise to short-run variable cost minimizes long-run cost at the given output vector.

investments in unconsolidated subsidiaries, intangibles, and other assets; and y_5, the credit equivalent amount of off-balance-sheet activities.[6] The six inputs are: x_1, labor; x_2, physical capital; x_3, time deposits exceeding $100,000 (uninsured);[7] x_4, all other deposits (including insured deposits); x_5, all other borrowed funds, including foreign deposits, federal funds purchased, reverse repos, trading account liabilities, mandatory convertible securities, mortgage indebtedness, commercial paper, and all other borrowed funds; and k, equity capital consisting of equity, subordinated debt, and loan loss reserves. Except for equity capital, the other five input prices are computed as the expenditure on the input divided by the quantity of the input. The price of a dollar of after-tax profit in terms of before-tax dollars is $p_\pi = 1 / (1 - t)$, where the tax rate, t, is the highest marginal corporate tax rate in the state in which the bank holding company is headquartered plus the highest federal marginal tax rate (which is 35 percent). Revenue, $\boldsymbol{p} \cdot \boldsymbol{y} + m$, is the sum of interest and noninterest income.

We proxy *ex post* asset quality by the amount of nonperforming loans, which is the sum of past due loans, leases, and other assets, and assets in nonaccrual status, plus gross charge-offs, plus other real estate owned in satisfaction of debts (i.e., real estate owned due to foreclosures). We proxy *ex ante* asset quality by the average contractual interest rate, p_i, on the ith output. The difference between this yield and the risk-free rate captures the risk premium incurred by the asset. Thus, the contractual interest rate captures both a component of revenue and a dimension of asset quality. Since interest income is not reported for all the outputs we specify, we use the weighted average of output prices, $\tilde{p}$, which is measured as the ratio of interest income from accruing assets to the sum of all the outputs.

Table 1 describes the full sample we used in the estimation (dropping outliers eliminated 10 bank holding companies, yielding 842 firms for the estimations). Banks range in assets from $72 million to $2.19 trillion. Because of the flexible nature of the production model and the fact that we are controlling

---

[6] Some studies proxy the amount of off-balance-sheet activities by the net income they generate. However, this measure is biased downward by losses. The credit-equivalent amount is calculated by converting the various measures of off-balance-sheet activities into the equivalent amount of on-balance-sheet assets, adjusted by the latter's risk weight. Loans are weighted at 100 percent. A stand-by letter of credit is weighted at 100 percent, too, on the grounds that it generates the same amount of exposure to default risk as an on-balance-sheet loan.

[7] The limit was temporarily increased to $250,000 in October 2008 and permanently increased by the Dodd-Frank Act of 2010. In 2007 the limit was $100,000.

for risk preferences and asset quality by including a measure of nonperforming loans and the average contractual interest rate on output, the model permits including a wide range of bank sizes. Tables 2-5 partition the data by asset size in order to show how the variables in Table 1 differ from small to very large banks. There is no official definition of too big to fail, but asset size of $100 billion or more has been considered a threshold for too big to fail in some studies.[8]

As shown in Table 2, the mean level of loans as a proportion of total assets falls somewhat as banks get larger, from about 0.72 in the smallest category with assets under $0.8 billion to 0.58 in the largest category with assets over $100 billion. The liquid assets ratio in the three smaller groups is approximately 0.04 and, in the larger two groups, with assets over $50 billion, it is 0.08 to 0.09. Trading and other assets as a proportion of total assets rise with bank size, from 0.04 in the smallest group to 0.18 in the largest. The ratio of the credit-equivalent amount of off-balance-sheet assets to total assets also rises with bank size, from 0.03 to 0.66.

Table 3 details differences in input utilization. While labor as a proportion of total assets does not vary much across the size groups, the physical capital ratio declines. Compared to smaller banks, larger banks fund a smaller proportion of their assets with insured deposits and a larger proportion with other borrowed funds (which include foreign deposits, commercial paper, federal funds purchased, securities sold under agreement to repurchase, trading account liabilities, and other borrowed money). Compared with insured deposits and other borrowed funds, uninsured deposits are a less important source of funds for all size groups.

Table 4 provides details of differences across size groups in risk exposure and financial performance. As banks increase in size across the six groups, their mean ratio of capital to assets increases from 0.099 to 0.13, while the mean ratio of nonperforming assets shows no monotonic pattern related to asset size. The rate of return on assets (ROA) (measured as profits/assets) in the largest group

[8] Brewer and Jagtiani (2009) give three too-big-to-fail size thresholds: (1) banks with total book value of assets of at least $100 billion, (2) banks that are one of the 11 largest organizations in each year (currently the 11th largest BHC has $290 billion in assets), and (3) banks with market value of equity ≥ $20 billion.

exceeds that of the other groups, but the differences in their mean ROA are negligible. The average contractual return on accruing assets is higher for smaller banks than for larger banks.

VIII. Evidence of Scale Economies

We estimate the cost function and input share equations for the theoretically mis-specified cash-flow cost function (omitting the amount of equity capital), the theoretically proper cash-flow cost function (conditioned on the amount of equity capital), and the most preferred profit function and input demand functions.

Table 6 presents the estimated scale economies for these models. The first column of results shows that for the mis-specified cost function that omits any role for equity capital, all six size groups show evidence of scale economies that are statistically significantly greater than one. The differences across size groups in these measured scale economies are slight. And the means differ little from the medians. To obtain some intuition for the magnitude of the measures, consider two values, 1.01 and 1.03 at each end of the range for the six groups. If all outputs increase by 10 percent, at a scale measure of 1.01, cost increases by 9.9 percent; and, at 1.03, cost increases by 9.7 percent.

The second column of results in Table 6 shows estimates of scale economies for the theoretically correct specification that includes equity capital as a conditioning argument but omits the cost of equity in the calculation of cost. Hence, we term the result "cash-flow cost." For banks in the four size groups with less than \$50 billion in total assets, we find essentially constant returns to scale, with measures of scale economies between 0.94 and 0.96. For banks larger than \$50 billion, we find diseconomies of scale, significant at the 1 percent level, with the mean estimate of scale economies at 0.90 and 0.88 in the two largest size categories. (These values imply that a 10 percent increase in all outputs results in an 11 percent increase in cost.) Both of the cash-flow cost functions in columns 1 and 2 omit the cost of equity in the calculation of cost used in the estimation. However, the cost function reported in the second column controls for the quantity of equity capital. Table 5 shows that larger banks, on average, fund their assets with relatively more equity than smaller banks. The additional equity used by larger banks to fund a dollar of assets reduces the amount of debt needed to fund that dollar of assets and, thus, reduces the relative interest expense of larger banks. Failing to control for the level of equity capital upwardly biases the estimated cost elasticity of larger banks in the mis-specified cost function in column 1. The cash-flow cost function in column 2 controls for this effect and returns a larger cost elasticity (i.e., smaller measure of scale economies) for all banks and considerably increased cost elasticity for banks with over $50 billion in assets – indeed, these banks show evidence of significant scale *diseconomies*.

Both cash-flow cost functions omit the cost of equity capital in their measure of cost and, hence, in their estimated cost elasticity. Column 3 of Table 6 reports these scale economies based on economic cost that includes the cost of equity. Adding the cost of equity increases the scale economies for all size categories. Banks with less than $50 billion in assets experience scale economies between 1.03 and 1.04, with the values significant at the 1 percent level for banks in the smaller size categories. Banks with total assets greater than $100 billion experience constant returns to scale.[9]

The failure of these properly specified minimum cost functions to show evidence of scale economies at larger banks may follow from their inability to distinguish the differences in risk-expected-return trade-offs that are inherent in the investment strategies of large and small banks. The most preferred cost function controls for these differences. We report the estimates of scale economies obtained from this cost function in column 4. The mean value of scale economies for the full sample is a significant 1.149. The categories of banks with assets less than $50 billion experience mean scale economies in the range 1.13 to 1.18.[10] For banks with assets between $50 billion and $100 billion, mean scale economies are 1.23, while for banks with assets over $100 billion mean scale economies are 1.35. The median values are more modest: 1.20 and 1.25, respectively.[11]

---

[9] At a value of 1.03, a 10 percent increase in all outputs would imply a 9.7 percent increase in cost; at 1.04, a 9.6 percent increase in cost; and at 1.00, a 10 percent increase in cost.

[10] In the case of 1.13, a 10 percent increase in all outputs would imply an 8.8 percent increase in cost, and for 1.18, an 8.5 percent increase in cost.

[11] A 10 percent increase in all outputs at 1.23 implies an increase in cost of 8.1 percent, while 1.35 implies a 7.4 percent increase in cost.

Robustness. Even though our model is very flexible and we control for risk preferences and output quality, there may be some concern that we are including banks with very different production technologies in the estimation and that this is driving our results. However, this does not appear to be the case. First, we re-estimate our model excluding banks with assets of $2 billion or less. This leaves a sample of 215 bank holding companies. Our scale results are very similar to those obtained with the full sample. Scale economies are significantly different from one at the 1 percent level and increase with bank size, from 1.14 for banks in the $2 billion to $10 billion size category up to 1.29 for banks with assets greater than $100 billion. Our results are also robust to re-estimating the model for the sample of bank that omits those with extreme values of output shares. This leaves a sample of 833 bank holding companies. Scale economies are significantly different from one at the 1 percent level and increase with bank size, from 1.15 for banks with assets under $0.8 billion and 1.30 for banks with assets greater than $100 billion.

IX. Evidence on Whether Scale Economies Are Driven by Too-Big-To-Fail Considerations

One question is whether the scale economies we find at very large banks are driven by their being too big to fail (TBTF), which might give them a cost advantage over other banks. There is no simple categorization of banks as TBTF. For the purposes of our analysis, let's consider banks with assets greater than $100 billion as being TBTF, which is consistent with the definitions suggested in Brewer and Jagtiani (2009). Here we present evidence that our scale results are not driven solely by TBTF considerations.

First, as presented above, we find scale economies not only at banks with assets > $100 billion but also at smaller banks, which are too small to be considered TBTF under any reasonable definition.

Second, we re-estimated our cost model for our sample of banks dropping the TBTF banks, i.e, banks with assets > $100 billion, and then calculated what scale economies would be for the TBTF banks, and for banks of other sizes, using this parameterization. Here, we once again found significant scale economies that increase with bank size. For banks with assets > $100 billion, the mean scale economies were 1.38 (compared with 1.35 in the baseline model estimated with the full sample of banks discussed above).

Third, to the extent that TBTF enables banks to enjoy lower funding costs because of lower risk premiums on the borrowed funds, it could be that our finding of scale economies at the largest banks in the sample is driven by these lower funding costs and that if these banks faced the same cost of funds as smaller banks, they would not enjoy scale economies. To investigate this possibility, we calculated what the scale economies for the TBTF banks would have been had the cost of the three inputs representing funding costs, namely, w_3 = uninsured deposit rate, w_4 = insured deposit rate, and w_5 = other borrowed funds rate, been the median values for the banks with assets ≤ \$100 billion. Again, we find significant scale economies that also increase with size. For banks with assets > \$100 billion, the mean scale economies were 1.37 (compared with 1.35 in the baseline model).

Thus, while there may be a funding cost advantage among the largest banks (perhaps because they are considered TBTF), our production model controls for this funding advantage in its computation of scale economies and there is no evidence that a funding cost advantage influences scale economies.

X. Policy Implications

A current policy question is how regulators should handle TBTF banks. One suggestion has been to impose a size limit on banks to try to prevent them from growing to be too big to fail in the first place. As discussed in Mester (2010), there would be several consequences of such a size limit, some of which might be unintended. Indeed, should scale economies be as strong as suggested in our results, banks would be motivated to try to circumvent such a limit. On the face of it, our estimates of scale economies suggest that such a size limit, by limiting the attainment of scale economies, would be quite costly. However, this is actually a more difficult question than it might seem. Typically, when researchers perform such calculations, they vary the scale of operations alone. And the estimates of scale economies essentially do that as well, by keeping product mix constant as scale is expanded. However, not only is the scale of operations different for large and small banks, the output mix also differs considerably, e.g., large banks have a considerably higher share of off-balance-sheet output. This variation in output mix turns out to be important when evaluating the potential cost impact of a size limit on banks.

In particular, we ask, what would be the change in cost if we broke up the 17 banks with assets greater than \$100 billion into banks with assets of \$100 billion. We will decompose the change in costs into two parts: the *scale effect*, which calculates the change in costs ignoring any change in output mix, and the *mix effect*, which calculates the change in costs from the change in the output mix that occurs when scale changes. Let YL = total assets of a bank with assets > \$100 billion (a large bank), YS = the size limit we are imposing (here, \$100 billion), HL represent the output mix (output shares) of the large bank, and HS represent the output mix of a \$100 billion bank. Then based on our estimated cost function, we can compute the ratio of the estimated cost of a set of n \$100 billion banks, $nC(YS,HS)$, to the estimated cost of a large bank, $C(YL,HL)$, where $n = YL/YS$:

$$(15)\qquad \frac{nC(YS,HS)}{C(YL,HL)} = \frac{\frac{YL}{YS}C(YS,HS)}{C(YL,HL)} = \frac{\frac{YL}{YS}C(YS,HL)}{C(YL,HL)} \times \frac{\frac{YL}{YS}C(YS,HS)}{\frac{YL}{YS}C(YS,HL)}$$

$$= \quad \text{scale effect} \times \text{mix effect.}$$

Our estimated scale economies can be used to calculate the *scale effect*:

$$(16)\qquad \text{scale effect} = \frac{\frac{YL}{YS}C(YS,HL)}{C(YL,HL)} = \frac{YL}{YS}\left\{\left[\left(\frac{YL}{YS}-1\right)\frac{1}{scale}\right]+1\right\}.$$

To calculate the *mix effect*, we need to know what product mix a bank with \$100 billion in assets would produce in order to calculate $C(YS,HS)$. Because there is no bank in the sample that has \$100 billion in assets (and even if there were, it would not necessarily be representative), we calculate the *mix effect* in two different ways. First, we approximate $C(YS,HS)$ by the mean $C(\cdot)$ of the 10 firms in the size category \$60 billion to \$140 billion in assets, which spans \$100 billion. Second, we approximate $C(YS,HS)$ by evaluating the estimated cost function *C(·)* at the median output shares and other non-output variables of banks in the asset size category of \$50 billion to \$100 billion, where we adjust y_3 so that the output levels sum to \$100 billion.

Our results indicate that breaking up the 17 banks in our sample with assets > \$100 billion into smaller banks with \$100 billion in assets but with no change in output mix would increase costs: the scale effect implies costs would increase by an estimated \$990 billion, or 2.4 times the sum of estimated costs of the 17 banks. This is consistent with the large scale economies we found. However, banks that are forced to downsize might also change their product mix over the longer run to one more consistent with a smaller scale of operations. Our estimated mix effect suggests that adjusting their output shares to those appropriate to their smaller size would lower costs by \$1.0-\$1.1 trillion. Thus, on net, the total longer-run impact of breaking up the banks into smaller institutions would be a cost savings of \$47-\$147 billion.

These calculations are only intended to be suggestive of one issue that must be considered in calculating the cost impact of imposing a size limit, namely, the effect on costs not only of a change in the scale of operations but also in the mix of outputs banks would choose to produce. Of course, these calculations ignore other consequences of such a policy. Moreover, they are only rough estimates and are dependent on the method of calculation. To see this, notice that another method of calculation indicates much smaller cost savings: Based on our estimates, the sum of estimated costs for the 17 banks in the largest size category with assets > \$100 billion is \$406 billion, and these banks hold a total of \$9.1 trillion in assets. The sum of estimated costs for the 12 banks in the second largest size category with assets between \$50 billion and \$100 billion is \$33 billion, and these banks hold a total of \$778 billion in assets. A simple back-of-the-envelope calculation indicates that redistributing the \$9.1 trillion of assets in the largest size category to the next largest size category would result in costs of \$385 billion (= (9.1 trillion / 778 billion) * \$33 billion), which, compared to the \$406 billion cost of banks in the largest category, suggests a cost savings of \$21 billion. Again, such a calculation assumes a change in output mix to that of banks in the second largest size category (and it also assumes that the other variables in the cost function, in particular, input prices, would be those consistent with those at banks in the second largest size category). The results again suggest that the effect of the change in output mix on costs dominates the cost savings attendant to increased scale via scale economies.

XI. Conclusions

We find evidence of large scale economies at smaller banks and even larger economies at large banks – economies consistent with the standard textbook arguments – when the production model endogenizes managers' choice of risk vs. expected return. The standard minimum cost function, even one that controls for equity capital, is not able to capture these scale economies.

Our results indicate that these measured scale economies do not result from cost advantages large banks may derive from too-big-to-fail considerations. Instead, they follow from technological advantages, such as diversification and the spreading of information costs and other costs that do not increase proportionately with size. Significant scale economies in banking suggest that technological factors, as well as TBTF cost advantages, appear to have been an important driver of banks' increasing size. While we do not know if the benefits of large size outweigh the potential costs in terms of systemic risk that large scale may impose on the financial system, our results suggest that strict size limits to control such costs will not likely be effective, since they work against market forces. Our results also indicate that one should consider both scale and product mix when evaluating such a policy.

Appendix

Empirical model and estimation: The Almost Ideal Production System[12]

The managers' most preferred (MP) production model comprises the profit share equation (3a), the input share equations (3b), and the first-order condition for the optimal level of equity capital, *k*, which is a conditioning argument in the share equations. The profit and input demand functions are shares expressed as shares of total revenue, $\boldsymbol{p} \cdot \boldsymbol{y} + m$, and sum to one. They are derived by applying Shephard's Lemma to the managerial expenditure function, which is dual to the utility maximization problem (1a-1c). Thus, the model to be estimated is:

(A1.1)
$$\frac{p_\pi \pi}{\mathbf{p}\cdot\mathbf{y}+m} = \frac{\partial \ln \boldsymbol{P}}{\partial \ln p_\pi} + \mu[ln(\mathbf{p}\cdot\mathbf{y}+m) - \ln \boldsymbol{P}]$$

(A1.2)
$$\frac{w_i x_i}{\mathbf{p}\cdot\mathbf{y}+m} = \frac{\partial \ln \boldsymbol{P}}{\partial \ln w_i} + \nu_i[ln(\mathbf{p}\cdot\mathbf{y}+m) - \ln \boldsymbol{P}]\ \forall i$$

(A1.3)
$$\frac{\partial V(\cdot)}{\partial k} = \frac{\partial V(\cdot)}{\partial \ln k}\frac{\partial \ln k}{\partial k} = 0,$$

where $ln\ \mathbf{P} = A_0 + \Sigma_i A_i\ ln\ y_i + (½)\Sigma_i \Sigma_j S_{ij}\ ln\ y_i\ ln\ y_j + \Sigma_i B_i\ ln\ w_i + (½)\Sigma_i \Sigma_j G_{ij}\ ln\ w_i\ ln\ w_j$

$+ \Sigma_i \Sigma_j D_i\ ln\ y_i\ ln\ w_j + (½)\Sigma_i \Sigma_j R_{ij}\ ln\ z_i\ ln\ z_j + \Sigma_i \Sigma_j H_{ij}\ ln\ z_i\ ln\ y_j + \Sigma_i \Sigma_j T_{ij}\ ln\ z_i\ ln\ w_j$;

$\mathbf{z} = (k, n, \boldsymbol{p}, p_\pi)$, and

(A1.4)
$$V(\cdot) = \frac{\ln(\mathbf{p}\cdot\mathbf{y}+m) - \ln \boldsymbol{P}}{\beta_0\left(\prod_i y_i^{\beta_i}\right)\left(\prod_j w_i^{\nu_i}\right)p_\pi{}^{\mu}k^{\kappa}}.$$

To save on degrees of freedom, in our estimation we replace the vector of output prices, ***p***, with the weighted-average output price, $\tilde{p} = \Sigma_i p_i (y_i/\Sigma_j y_j)$. The risk-free rate, *r*, is the same for all bank

[12] The exposition in this appendix is adapted from Hughes, Lang, Mester, and Moon (2000).

holding companies, so the coefficients on terms involving r are not estimated. Written out, the equations to be estimated are:

$$\begin{aligned}\frac{p_\pi \pi}{\mathbf{p}\cdot\mathbf{y}+m} &= F_4 + R_{44}\ln p_\pi + R_{34}\ln\tilde{p} + \sum_j H_{4j}\ln y_j + \sum_s T_{4s}\ln w_s \\ &\quad + R_{24}\ln n + R_{14}\ln k + \mu[ln(\mathbf{p}\cdot\mathbf{y}+m) - \ln \boldsymbol{P}] \\ &= F_4 + \sum_j H_{4j}\ln y_j + \sum_s T_{4s}\ln w_s \\ &\quad + \sum_i R_{i4}\ln z_i + \mu[ln(\mathbf{p}\cdot\mathbf{y}+m) - \ln \boldsymbol{P}]\end{aligned} \tag{A1.1$'$}$$

$$\begin{aligned}\frac{w_i x_i}{\mathbf{p}\cdot\mathbf{y}+m} &= B_i + \sum_s G_{ij}\ln w_s + T_{3i}\ln\tilde{p} + \sum_j D_{ji}\ln y_j + T_{4i}\ln p_\pi \\ &\quad + T_{3i}\ln n + T_{1i}\ln k + \nu_i[ln(\mathbf{p}\cdot\mathbf{y}+m) - \ln \boldsymbol{P}] \\ &= B_i + \sum_j D_{ji}\ln y_j + \sum_s G_{ij}\ln w_s \\ &\quad + \sum_s T_{si}\ln z_s + \nu_i[ln(\mathbf{p}\cdot\mathbf{y}+m) - \ln \boldsymbol{P}]\end{aligned} \tag{A1.2$'$}$$

$$\begin{aligned}&F_1 + R_{11}\ln k + R_{13}\ln\tilde{p} + \sum_j H_{1j}\ln y_j + \sum_s T_{1s}\ln w_s + R_{14}\ln p_\pi \\ &\quad + R_{12}\ln k + \kappa[ln(\mathbf{p}\cdot\mathbf{y}+m) - \ln \boldsymbol{P}] = 0 \\ &\Rightarrow F_1 + \sum_j H_{1j}\ln y_j + \sum_s T_{1s}\ln w_s + \sum_j R_{1j}\ln z_j + \kappa[ln(\mathbf{p}\cdot\mathbf{y}+m) - \ln \boldsymbol{P}] = 0\end{aligned} \tag{A1.3$'$}$$

where

$$\begin{aligned}
\ln \boldsymbol{P} = {} & A_0 + F_3 \ln \tilde{p} + \sum_i A_i \ln y_i + \sum_j B_j \ln w_j \\
& + F_4 \ln p_\pi + F_2 \ln n + F_1 \ln k + \frac{1}{2} R_{33} (\ln \tilde{p})^2 + \frac{1}{2} \sum_i \sum_j S_{ij} \ln y_i \ln y_j \\
& + \frac{1}{2} \sum_s \sum_t G_{st} \ln w_s \ln w_t + \frac{1}{2} R_{44} (\ln p_\pi)^2 \\
& + \frac{1}{2} R_{22} (\ln n)^2 + \frac{1}{2} R_{11} (\ln k)^2 \\
& + \sum_j H_{3j} \ln \tilde{p} \ln y_j + \sum_s T_{3s} \ln \tilde{p} \ln w_s + R_{34} \ln \tilde{p} \ln p_\pi \\
& + R_{23} \ln \tilde{p} \ln n + R_{13} \ln \tilde{p} \ln k \\
& + \sum_i \sum_s D_{is} \ln y_i \ln w_s + \sum_j H_{4j} \ln y_j \ln p_\pi \\
& + \sum_j H_{2j} \ln y_j \ln n + \sum_j H_{1j} \ln y_j \ln k \\
& + \sum_s T_{4s} \ln w_s \ln p_\pi \\
& + \sum_s T_{2s} \ln w_s \ln n + \sum_s T_{1s} \ln w_s \ln k \\
& + R_{24} \ln p_\pi \ln n + R_{14} \ln p_\pi \ln k \\
& + R_{12} \ln n \ln k
\end{aligned}$$

$$\begin{aligned}
& = A_0 + \sum_i A_i \ln y_i + \sum_s B_s \ln w_s + \sum_j F_j \ln z_j \\
& + \frac{1}{2} \sum_i \sum_j S_{ij} \ln y_i \ln y_j + \frac{1}{2} \sum_s \sum_t G_{st} \ln w_s \ln w_t + \frac{1}{2} \sum_i \sum_j R_{ij} \ln z_i \ln z_j \\
& + \sum_i \sum_s D_{is} \ln y_i \ln w_s + \sum_i \sum_j H_{ij} \ln z_i \ln y_j + \sum_i \sum_s T_{is} \ln z_i \ln w_s
\end{aligned}$$

and $\quad p_\pi = 1/(1-t)$.

We impose several conditions on the parameters of the model. *Symmetry* requires that[13]

(S1) $S_{ij} = S_{ji} \ \forall$ i,j,

(S2) $T_{4s} = T_{s4} \ \forall$ s, and

(S3) $G_{si} = G_{is} \ \forall$ s,i.

[13] (S1) must be imposed in the estimation of the share equations, since the constituent coefficients cannot be separately identified. However, (S2) and (S3) involve coefficients of prices that are used by Shephard's Lemma to obtain the share equations. Consequently, they appear in separate share equations and are, thus, identifiable. It is a judgment call as to whether one imposes these symmetry conditions. We impose them in our estimation.

The input and profit revenue share equations sum to one, which implies the following *adding-up* conditions:

(A1) $\Sigma_i B_i + F_4 = 1$,

(A2) $\Sigma_i G_{si} + T_{4s} = 0, \forall s$,

(A3) $\Sigma_i T_{3i} + R_{34} = 0$,

(A4) $\Sigma_i D_{ji} + H_{4j} = 0 \ \forall j$,

(A5) $\Sigma_i T_{4i} + R_{44} = 0$,

(A6) $\Sigma_i T_{1i} + R_{14} = 0$,

(A7) $\Sigma_i T_{2i} + R_{24} = 0$, and

(A8) $\Sigma v_j + \mu = 0$.

The input and profit share equations are *homogeneous of degree zero* in ($\boldsymbol{w}$, $\tilde{p}$, r, p_π). The only homogeneity condition imposed in the estimation is:

(H1) $\Sigma v_j + \mu = 0$,

which is equivalent to the adding-up condition (A8). The other homogeneity conditions contain coefficients on variables involving the risk-free rate, *r*. These coefficients are not estimated, since *r* does not vary across banks, but the homogeneity conditions can be used to recover these coefficients.
To summarize: in estimating the model, we imposed (S1), (A1)-(A7), and (A8) ≡ (H1).

We estimated the model using nonlinear two-stage least squares, a generalized method of moments. Starting values were obtained by setting the constant terms, B_i, in the input share equations at the average value of the input share across banks in the sample, the constant term, *F4*, in the profit share equation at the average value of the profit share across banks in the sample, and all other parameters in the input share, profit share, and equity capital demand equation equal to 0.

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* **21**, 895-947.

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

Braeutigam, R. R. and Daughety, A. F. (1983). On the estimation of returns to scale using variable cost functions, *Economic Letters* **11**, 25-31.

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.

Deaton, A. and Muellbauer, J. (1980). An almost ideal demand system, *American Economic Review* **70**, 312-326.

Feng, G. and A. Serletis. (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* **34**, 127-138.

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

Hughes, J.P., Lang, W., Mester, L.J. and Moon C.-G. (1996). Efficient banking under interstate branching, *Journal of Money, Credit, and Banking* **28**, 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* **18**, 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* **80**, 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., 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* **25**, 2169-2208.

Keeley, Michael C. (1990). Deposit insurance, risk, and market power in banking, *American Economic Review* **80**, 1183-1200.

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

Mester, L.J. (2010). Scale economies in banking and financial regulatory reform, *The Region*, Federal Reserve Bank of Minneapolis, September 2010, 10-13.

Tufano, Peter (1996). Who manages risk? An empirical examination of risk management practices in the gold mining industry, *Journal of Finance* **50**, 1097-1137.

Wheelock, D. C. and P. W. Wilson. (2009). Are U. S. banks too large? Working Paper 2009-054B, Federal Reserve Bank of St. Louis, December.

Figure 1

A smaller bank's investment strategies are depicted on the risk-return frontier labeled *I* and the larger bank's strategies on frontier *II*. The improved trade-off along frontier *II* results from the better diversification of the larger bank. Point *A* represents production of a smaller, less diversified output, say, some quantity of loans with a particular probability distribution of default that reflects the contractual interest rate charged and the resources allocated to risk assessment and monitoring. Point *B* represents a larger quantity of loans with the same contractual interest rate but better diversification and, hence, an improved probability distribution of default and lower overall risk. The better diversification allows the costs of risk management to increase less than proportionately with the loan volume while maintaining an improved probability distribution of default. Thus, the response of cost to the increase in output from point *A* to point *B* reflects economies of scale. And the expected return at *B* exceeds that at *A*.

On the other hand, suppose the bank responds to the better diversification of the larger output by adopting a more risky investment strategy for an enhanced expected return. Better diversification does not offset the increased cost occasioned by the additional default risk. Point *D* in Figure 1 designates this strategy. The increased inherent default risk due to the higher contractual interest rate results in costs of risk management that increase more than proportionately with the loan volume (from *A* to *D*), and production appears to exhibit the counter-intuitive *scale diseconomies* found by empirical studies of banking cost that fail to account for endogenous risk-taking.

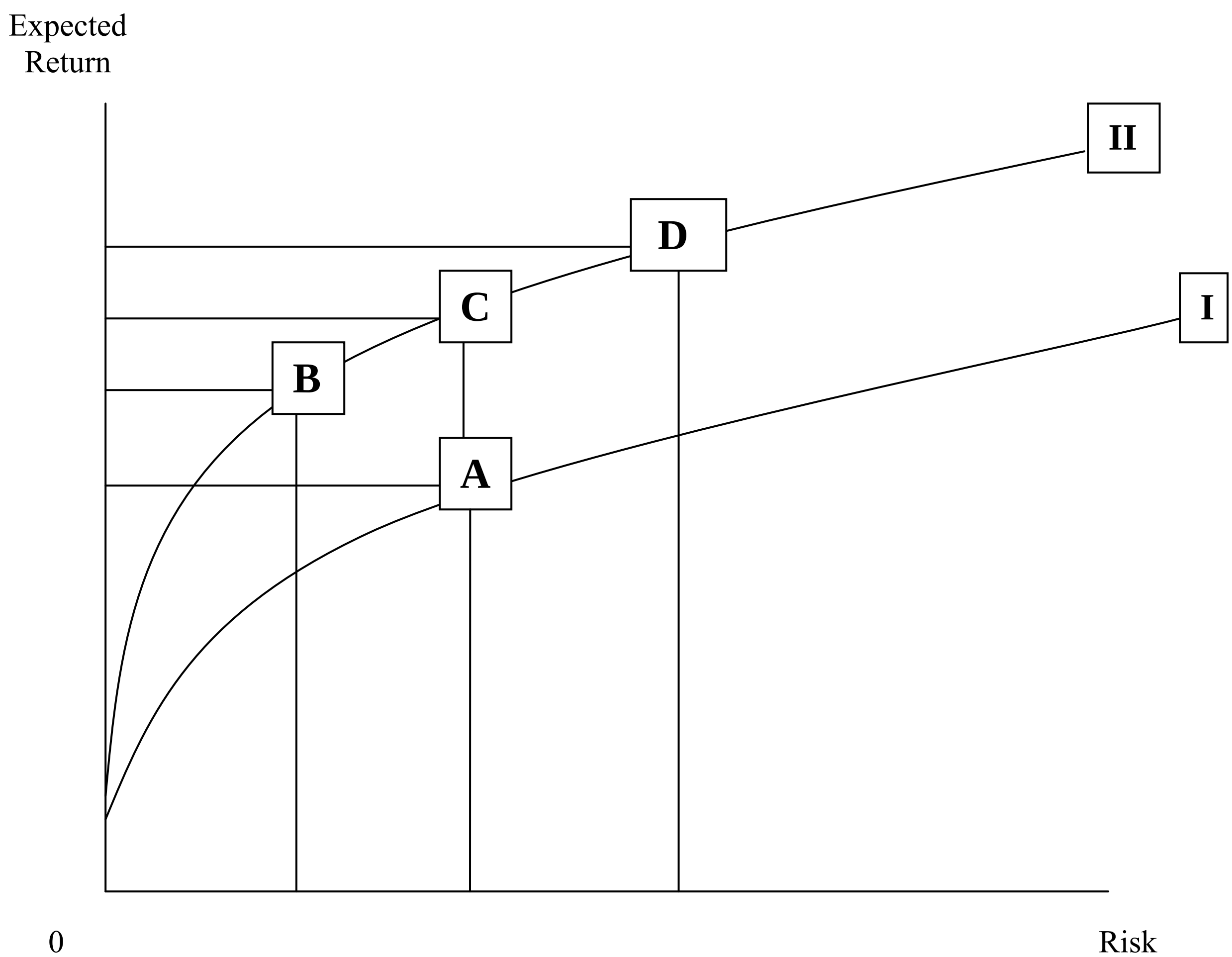

Figure 2

The investment strategies in Figure 1 are illustrated for the case of the larger output along frontier *II*. The production technology for a given quantity of loans is represented by the isoquant. The mix of debt and equity used to fund the loans is ignored. The diagram shows the quantity of physical capital and labor used in the process of credit evaluation and loan monitoring. Point *C* shows the least costly way to produce the particular quantity of loans with the risk exposure associated with the investment strategy *C* in Figure 1. If a bank adopted the less risky strategy, *B*, it might use less labor in credit evaluation and monitoring: point *B* in Figure 2, a less costly method of producing the *same quantity of loans*. Thus, the isoquant for this quantity of loans that passes through point *C* captures one investment strategy only. If the isoquant included a characterization of the risk exposure, there would be another isoquant passing through point *B* for the same quantity of loans produced with the lower risk strategy. On the other hand, if a bank adopted the more risky strategy, *D*, it would use more labor, the corresponding point *D* in Figure 2, a more costly method than *C*. Thus, the cost of producing this particular quantity of loans depends on a bank's choice of risk exposure and its expected return. We shall refer to this characterization of cost as *risk-return-driven cost*.

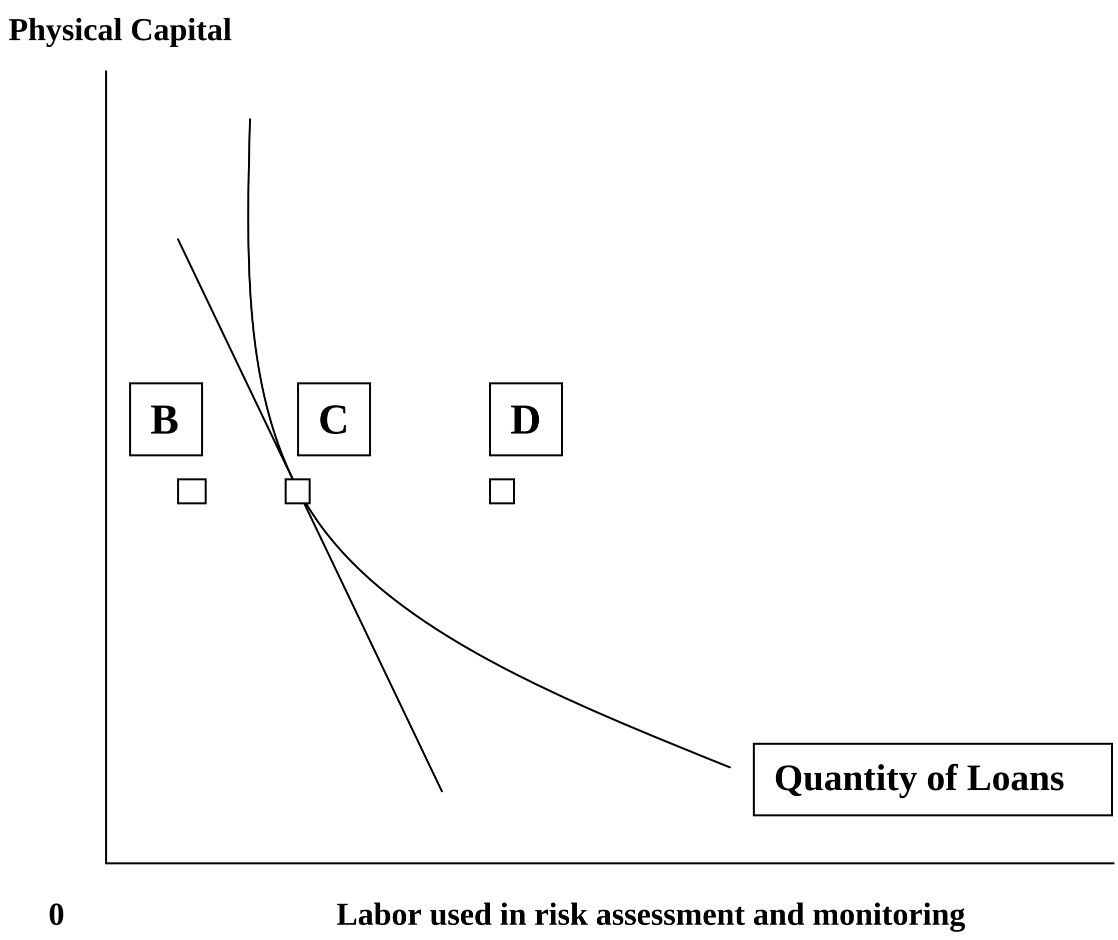

Table 1. Summary Statistics: Full Sample

The data, obtained from the Y9-C Call Reports filed quarterly with regulators, include 842 top-tier U.S. bank holding companies in 2007. A top-tier company is not owned by another company.

Variable	Mean	Median	Std Dev	Minimum	Maximum
Total Assets in $1000s	13,692,833	941,224	117,267,435	72,238	2,187,631,000
Total Revenue in $1000s	1,024,870	70,026	8,568,377	6,555	158,039,000
Financial Performance					
Equity Capital/Assets	0.102	0.097	0.026	0.043	0.357
Nonperforming Assets/Assets	0.022	0.016	0.018	0.000	0.243
Profit/Revenue	0.317	0.315	0.079	-0.299	0.613
Profit/Assets	0.024	0.023	0.010	-0.014	0.096
Asset Allocation					
Liquid Assets: y_1/Assets	0.044	0.033	0.046	0.003	0.429
Securities: y_2/Assets	0.174	0.159	0.100	0.000	0.567
Loans: y_3/Assets	0.718	0.734	0.116	0.111	0.949
Trading, Other Assets: y_4/Assets	0.051	0.041	0.036	0.008	0.409
Off-Balance-Sheet Items: y_5/Assets	0.060	0.034	0.226	0.000	4.142
Input Utilization					
Labor (FTEs): x_1/assets	0.00027	0.00027	0.00011	0.000031	0.0023
Physical Capital: x_2/Assets	0.019	0.018	0.010	0.001	0.061
Uninsured Deposits: x_3/Assets	0.149	0.134	0.073	0.006	0.552
Insured Deposits: x_4/Assets	0.610	0.625	0.114	0.079	0.854
Other Borrowed Funds: x_5/assets	0.122	0.104	0.109	0.000	0.716
Prices					
Average Interest Rate on Assets	0.061	0.061	0.011	0.007	0.095
Wage Rate: w_1 in $1000s	63.322	59.391	22.224	32.512	189.621
Price of Physical Capital: w_2	0.288	0.215	0.248	0.059	2.344
Uninsured Deposit Rate: w_3	0.048	0.047	0.011	0.008	0.119
Insured Deposit Rate: w_4	0.028	0.027	0.005	0.002	0.059
Other Borrowed Funds rate: w_5	0.054	0.048	0.020	0.013	0.179
Tax Rate	0.421	0.420	0.115	0.350	0.470
1/(1–Tax Rate)	1.730	1.724	0.241	1.538	1.887

Table 2. Summary Statistics: Asset Allocation by Size Groups

The data, obtained from the Y9-C Call Reports filed quarterly with regulators, include 842 top-tier U.S. bank holding companies in 2007. A top-tier company is not owned by another company. Banks in the largest size category, with assets exceeding $100 billion, are often perceived as being too big to fail (Brewer and Jagtiani, 2009).

Total Assets < $0.8 billion (n = 328)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Total Assets in $1000s	571,582.287	589,526.000	139,619.056	72,238.000	799,832.000
Liquid Assets: y1/Assets	0.045	0.036	0.031	0.005	0.318
Securities: y2/Assets	0.182	0.163	0.103	0.006	0.567
Loans: y3/Assets	0.719	0.739	0.108	0.401	0.915
Trading, Other Assets: y4/Assets	0.040	0.036	0.024	0.008	0.228
Off-Balance-Sheet Items: y5/Assets	0.032	0.025	0.029	0.000	0.181

Total Assets $0.8 billion – $2 billion (n = 299)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Total Assets in $1000s	1,199,358.877	1,123,172.000	334,356.263	800936.000	1,999,119.000
Liquid Assets: y1/Assets	0.041	0.032	0.038	0.007	0.400
Securities: y2/Assets	0.163	0.150	0.093	0.000	0.535
Loans: y3/Assets	0.738	0.753	0.102	0.388	0.949
Trading, Other Assets: y4/Assets	0.045	0.040	0.023	0.008	0.169
Off-Balance-Sheet Items: y5/Assets	0.041	0.032	0.032	0.001	0.190

Total Assets $2 billion – $10 billion (n = 155)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Total Assets in $1000s	4,091,767.774	3,350,126.000	2,031,203.785	2,012,453.000	9,731,046.000
Liquid Assets: y1/Assets	0.038	0.029	0.029	0.003	0.179
Securities: y2/Assets	0.177	0.164	0.098	0.003	0.551
Loans: y3/Assets	0.716	0.725	0.103	0.388	0.946
Trading, Other Assets: y4/Assets	0.060	0.056	0.030	0.011	0.189
Off-Balance-Sheet Items: y5/Assets	0.055	0.046	0.041	0.001	0.303

Total Assets $10 billion – $50 billion (n = 31)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Total Assets in $1000s	16,562,010.323	13,871,556.00	6,989,958.953	10,402,532.00	37,017,239.000
Liquid Assets: y1/Assets	0.041	0.029	0.029	0.011	0.141
Securities: y2/Assets	0.187	0.170	0.063	0.078	0.373
Loans: y3/Assets	0.684	0.693	0.075	0.496	0.816
Trading, Other Assets: y4/Assets	0.078	0.069	0.032	0.023	0.157
Off-Balance-Sheet Items: y5/Assets	0.089	0.073	0.062	0.008	0.295

Total Assets $50 billion – $100 billion (n = 12)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Total Assets in $1000s	64,794,778.917	61,460,925.50	12,501,001.639	52,947,444.00	99,567,393.000
Liquid Assets: y1/Assets	0.079	0.044	0.113	0.022	0.429
Securities: y2/Assets	0.139	0.128	0.054	0.075	0.280
Loans: y3/Assets	0.660	0.739	0.142	0.375	0.812
Trading, Other Assets: y4/Assets	0.121	0.112	0.060	0.060	0.242
Off-Balance-Sheet Items: y5/Assets	0.441	0.221	0.785	0.062	2.893

Total Assets > $100 billion (n = 17)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Total Assets in $1000s	532,904,914.00	179,573,933.0	653,325,466.24	110,961,509.0	2,187,631,000.0
Liquid Assets: y1/Assets	0.087	0.039	0.082	0.015	0.248
Securities: y2/Assets	0.153	0.131	0.110	0.052	0.522
Loans: y3/Assets	0.577	0.668	0.193	0.111	0.800
Trading, Other Assets: y4/Assets	0.181	0.154	0.084	0.104	0.393
Off-Balance-Sheet Items: y5/Assets	0.657	0.250	1.181	0.037	4.142

Table 3. Summary Statistics: Input Utilization by Size Groups

The data, obtained from the Y9-C Call Reports filed quarterly with regulators, include 842 top-tier U.S. bank holding companies in 2007. A top-tier company is not owned by another company. Banks in the largest size-category, with assets exceeding $100 billion, are often perceived as being too big to fail (Brewer and Jagtiani, 2009).

Total Assets < $0.8 billion (n = 328)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Labor (FTEs): x_1/assets	0.00030	0.00029	0.000086	0.000095	0.00065
Physical Capital: x_2/Assets	0.0213	0.0198	0.0104	0.0018	0.0577
Uninsured Deposits: x_3/Assets	0.1558	0.1395	0.0698	0.0063	0.5516
Insured Deposits: x_4/Assets	0.6344	0.6465	0.0894	0.2330	0.8172
Other Borrowed Funds: x_5/assets	0.0969	0.0885	0.0676	0.0007	0.3903

Total Assets $0.8 billion – $2 billion (n = 299)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Labor (FTEs): x_1/assets	0.00027	0.00026	0.00015	0.000031	0.0023
Physical Capital: x_2/Assets	0.0198	0.0195	0.0098	0.0006	0.0555
Uninsured Deposits: x_3/Assets	0.1508	0.1380	0.0704	0.0142	0.4223
Insured Deposits: x_4/Assets	0.6185	0.6240	0.0981	0.1105	0.8386
Other Borrowed Funds: x_5/assets	0.1143	0.0984	0.0810	0.0001	0.6949

Total Assets $2 billion – $10 billion (n = 155)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Labor (FTEs): x_1/assets	0.00024	0.00024	0.000077	0.000074	0.00048
Physical Capital: x_2/Assets	0.0173	0.0157	0.0104	0.0015	0.0608
Uninsured Deposits: x_3/Assets	0.1471	0.1221	0.0808	0.0443	0.3977
Insured Deposits: x_4/Assets	0.5909	0.6050	0.1067	0.1302	0.8332
Other Borrowed Funds: x_5/assets	0.1384	0.1276	0.0802	0.0083	0.4783

Total Assets $10 billion – $50 billion (n = 31)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Labor (FTEs): x_1/assets	0.00022	0.00022	0.000076	0.000086	0.00044
Physical Capital: x_2/Assets	0.0157	0.0123	0.0094	0.0055	0.0467
Uninsured Deposits: x_3/Assets	0.1209	0.0944	0.0626	0.0230	0.2792
Insured Deposits: x_4/Assets	0.5532	0.5642	0.1203	0.3207	0.8539
Other Borrowed Funds: x_5/assets	0.1931	0.1745	0.0932	0.0412	0.3985

Total Assets $50 billion – $100 billion (n = 12)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Labor (FTEs): x_1/assets	0.00017	0.00018	0.000036	0.000093	0.00022
Physical Capital: x_2/Assets	0.0092	0.0083	0.0032	0.0057	0.0170
Uninsured Deposits: x_3/Assets	0.1061	0.1013	0.0464	0.0430	0.1783
Insured Deposits: x_4/Assets	0.4749	0.4922	0.1295	0.1910	0.6795
Other Borrowed Funds: x_5/assets	0.2563	0.2216	0.1593	0.0498	0.6471

Total Assets > $100 billion (n = 17)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Labor (FTEs): x_1/assets	0.00018	0.00019	0.000047	0.000098	0.00028
Physical Capital: x_2/Assets	0.0097	0.0087	0.0045	0.0030	0.0191
Uninsured Deposits: x_3/Assets	0.0769	0.0692	0.0326	0.0238	0.1350
Insured Deposits: x_4/Assets	0.3799	0.4604	0.1712	0.0791	0.5734
Other Borrowed Funds: x_5/assets	0.3670	0.2881	0.1814	0.1870	0.7161

Table 4. Summary Statistics: Risk and Financial Performance by Size Groups

The data, obtained from the Y9-C Call Reports filed quarterly with regulators, include 842 top-tier U.S. bank holding companies in 2007. A top-tier company is not owned by another company. Banks in the largest size-category, with assets exceeding $100 billion, are often perceived as being too big to fail (Brewer and Jagtiani, 2009).

Total Assets < $0.8 billion (n = 328)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Equity Capital/Assets	0.099	0.095	0.028	0.045	0.235
Average Interest Rate on Assets	0.064	0.063	0.007	0.043	0.095
Nonperforming Assets/Assets	0.025	0.018	0.027	0.000	0.239
Total Revenue ($1000s)	42,232.863	42,302.000	11,720.371	6,555.000	90,682.000
Profit/Revenue	0.302	0.304	0.078	-0.299	0.544
Profit/Assets	0.023	0.022	0.007	-0.014	0.056

Total Assets $0.8 billion – $2 billion (n = 299)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Equity Capital/Assets	0.100	0.095	0.028	0.046	0.357
Average Interest Rate on Assets	0.062	0.061	0.007	0.040	0.088
Nonperforming Assets/Assets	0.020	0.015	0.021	0.001	0.243
Total Revenue ($1000s)	88,222.164	81,848.000	27,634.679	44,180.000	214,208.000
Profit/Revenue	0.311	0.309	0.075	0.001	0.613
Profit/Assets	0.023	0.022	0.009	0.000	0.096

Total Assets $2 billion – $10 billion (n = 155)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Equity Capital/Assets	0.105	0.103	0.027	0.043	0.230
Average Interest Rate on Assets	0.060	0.060	0.007	0.037	0.087
Nonperforming Assets/Assets	0.020	0.016	0.016	0.001	0.109
Total Revenue ($1000s)	30,0481.432	246,612.000	156,223.469	112,616.000	876,904.000
Profit/Revenue	0.336	0.329	0.068	0.119	0.527
Profit/Assets	0.025	0.024	0.007	0.006	0.055

Total Assets $10 billion – $50 billion (n = 31)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Equity Capital/Assets	0.114	0.112	0.023	0.048	0.158
Average Interest Rate on Assets	0.055	0.056	0.005	0.036	0.062
Nonperforming Assets/Assets	0.017	0.015	0.009	0.002	0.042
Total Revenue ($1000s)	1,209,522.323	1,032,973.000	554,743.096	641,771.000	3,144,098.000
Profit/Revenue	0.353	0.355	0.069	0.213	0.478
Profit/Assets	0.026	0.023	0.007	0.015	0.039

Total Assets $50 billion – $100 billion (n = 12)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Equity Capital/Assets	0.142	0.135	0.045	0.086	0.259
Average Interest Rate on Assets	0.041	0.045	0.012	0.010	0.048
Nonperforming Assets/Assets	0.016	0.014	0.008	0.004	0.033
Total Revenue ($1000s)	4,359,240.750	4,349,244.500	1,395,613.601	1,991,499.000	7,890,349.000
Profit/Revenue	0.385	0.389	0.061	0.246	0.511
Profit/Assets	0.026	0.028	0.006	0.013	0.033

Total Assets > $100 billion (n = 17)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Equity Capital/Assets	0.131	0.130	0.034	0.073	0.195
Average Interest Rate on Assets	0.042	0.045	0.017	0.007	0.067
Nonperforming Assets/Assets	0.024	0.020	0.017	0.000	0.068
Total Revenue ($1000s)	40,372,301.176	15,015,000.00	46,498,991.907	8,467,745.000	158,039,000.00
Profit/Revenue	0.404	0.398	0.077	0.260	0.602
Profit/Assets	0.033	0.031	0.013	0.019	0.076

Table 5. Summary Statistics: Prices by Size Groups

The data, obtained from the Y9-C Call Reports filed quarterly with regulators, include 842 top-tier U.S. bank holding companies in 2007. A top-tier company is not owned by another company. Banks in the largest size-category, with assets exceeding $100 billion, are often perceived as being too big to fail (Brewer and Jagtiani, 2009).

Total Assets < $0.8 billion (n = 328)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Wage Rate: w_1	58.495	56.801	12.434	33.389	122.689
Price of Physical Capital: w_2	0.261	0.200	0.216	0.059	2.181
Uninsured Deposit Rate: w_3	0.049	0.047	0.011	0.018	0.119
Insured Deposit Rate: w_4	0.028	0.028	0.006	0.013	0.050
Other Borrowed Funds rate: w_5	0.057	0.050	0.027	0.013	0.179
Tax Rate	0.420	0.420	0.025	0.350	0.470
Price of After-Tax Profit (1/(1–t))	1.728	1.724	0.073	1.538	1.887

Total Assets $0.8 billion – $2 billion (n = 299)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Wage Rate: w_1	63.261	59.208	16.784	34.810	167.079
Price of Physical Capital: w_2	0.276	0.206	0.235	0.079	2.344
Uninsured Deposit Rate: w_3	0.049	0.048	0.012	0.016	0.113
Insured Deposit Rate: w_4	0.029	0.028	0.007	0.008	0.059
Other Borrowed Funds rate: w_5	0.053	0.049	0.024	0.014	0.178
Tax Rate	0.421	0.420	0.025	0.350	0.470
Price of After-Tax Profit (1/(1—-t))	1.730	1.724	0.072	1.538	1.887

Total Assets $2 billion – $10 billion (n = 155)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Wage Rate: w_1	68.491	62.710	22.557	35.224	189.621
Price of Physical Capital: w_2	0.345	0.251	0.282	0.062	1.789
Uninsured Deposit Rate: w_3	0.047	0.047	0.010	0.009	0.092
Insured Deposit Rate: w_4	0.026	0.026	0.007	0.002	0.048
Other Borrowed Funds rate: w_5	0.052	0.046	0.019	0.021	0.151
Tax Rate	0.421	0.423	0.026	0.350	0.470
Price of After-Tax Profit (1/(1–t))	1.729	1.733	0.075	1.538	1.887

Total Assets $10 billion – $50 billion (n = 31)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Wage Rate: w_1	68.748	66.348	16.664	37.031	111.536
Price of Physical Capital: w_2	0.302	0.273	0.117	0.150	0.563
Uninsured Deposit Rate: w_3	0.054	0.050	0.017	0.034	0.118
Insured Deposit Rate: w_4	0.024	0.025	0.006	0.012	0.042
Other Borrowed Funds rate: w_5	0.047	0.045	0.016	0.029	0.101
Tax Rate	0.422	0.423	0.026	0.350	0.458
Price of After-Tax Profit (1/(1–t))	1.735	1.733	0.075	1.538	1.846

Total Assets $50 – $100 billion (n = 12)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Wage Rate: w_1	79.908	73.445	25.567	32.512	123.813
Price of Physical Capital: w_2	0.347	0.342	0.116	0.102	0.503
Uninsured Deposit Rate: w_3	0.046	0.047	0.008	0.029	0.055
Insured Deposit Rate: w_4	0.020	0.020	0.005	0.012	0.029
Other Borrowed Funds rate: w_5	0.039	0.044	0.010	0.017	0.050
Tax Rate	0.424	0.427	0.020	0.395	0.458
Price of After-Tax Profit (1/(1–t))	1.737	1.745	0.059	1.653	1.846

Total Assets > $100 billion (n = 17)

Variable	Mean	Median	Std Dev	Minimum	Maximum
Wage Rate: w_1	88.800	84.380	19.392	59.816	133.045
Price of Physical Capital: w_2	0.443	0.394	0.174	0.188	0.794
Uninsured Deposit Rate: w_3	0.043	0.048	0.016	0.008	0.066
Insured Deposit Rate: w_4	0.023	0.023	0.005	0.014	0.031
Other Borrowed Funds rate: w_5	0.043	0.043	0.008	0.028	0.054
Tax Rate	0.430	0.425	0.015	0.410	0.458
Price of After-Tax Profit (1/(1–t))	1.756	1.739	0.048	1.695	1.846

Table 6
Estimated Mean Scale Economies

Scale economies are calculated as the mean of the estimated scale economies at each point in the sample or size category (rather than scale economies evaluated at the mean of the data). The data, obtained from the Y9-C Call Reports filed quarterly with regulators, include 842 top-tier U.S. bank holding companies in 2007. A top-tier company is not owned by another company. Banks in the largest size-category, with assets exceeding $100 billion, are often perceived as being too big to fail (Brewer and Jagtiani, 2009).

The estimations include the cost function and input share equations for the theoretically mis-specified cash-flow cost function (omitting the amount of equity capital) and the theoretically proper cash-flow cost function (conditioned on the amount of equity capital). The economic-cost scale economies are inferred from the theoretically proper cash-flow cost function. In addition, we estimate the managers' most preferred profit function and input demand functions, which reflect the bank's risk-expected-return trade-off, and compute the managers' most-preferred cost function from the profit function.

	(1) Mis-specified Cash-Flow Cost Function		**(2) Correct Cash-Flow Cost Function**		**(3) Economic Cost Function**		**(4) Managers' Most Preferred Cost Function**	
	Omits Level of Equity		**Conditioned on Level of Equity**		**Includes Shadow Cost of Equity**		**Conditioned on Optimal Equity**	
Total Assets	**Mean**	**Median**	**Mean**	**Median**	**Mean**	**Median**	**Mean**	**Median**
Full sample n = 842	**1.0319** (0.0056)	1.0315	0.9542** (0.0190)	0.9561	**1.0389** (0.0083)	1.0386	**1.1490** (0.0095)	1.1341
< $0.8 billion n = 328	**1.0344** (0.0064)	1.0341	0.9606** (0.0201)	0.9610	**1.0430** (0.0093)	1.0422	**1.1364** (0.0087)	1.1272
$0.8 billion – $2 billion n = 299	**1.0328** (0.0057)	1.0319	0.9582** (0.0190)	0.9612	**1.0393** (0.0083)	1.0391	**1.1421** (0.0093)	1.1297
$2 billion – $10 billion n = 155	**1.0293** (0.0054)	1.0289	0.9490** (0.0212)	0.9522	**1.0331** (0.0093)	1.0328	**1.1549** (0.0103)	1.1470
$10 billion – $50 billion n = 31	**1.0246** (0.0064)	1.0255	0.9363** (0.0268)	0.9421	1.0265** (0.0126)	1.0258	**1.1782** (0.0135)	1.1518
$50 billion – $100 billion n = 12	**1.0211** (0.0077)	1.0212	**0.8977** (0.0360)	0.8981	1.0359** (0.0156)	1.0302	**1.2330** (0.0177)	1.1976
> $100 billion n = 17	1.0152 (0.0097)	1.0162	**0.8837** (0.0404)	0.8861	1.0279 (0.0193)	1.0215	**1.3478** (0.0295)	1.2508

Standard errors are given in parentheses.
All estimates of scale economies are significantly different from 0 at the 1 percent level.
Estimates of scale economies in **bold** are significantly different from 1 at the 1 percent level.
* Significantly different from 1 at the 10 percent level
** Significantly different from 1 at the 5 percent level