Wharton

Financial
Institutions
Center

Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?

by
Allen N. Berger
Loretta J. Mester

97-04

The Wharton School
University of Pennsylvania

THE WHARTON FINANCIAL INSTITUTIONS CENTER

The Wharton Financial Institutions Center provides a multi-disciplinary research approach to the problems and opportunities facing the financial services industry in its search for competitive excellence. The Center's research focuses on the issues related to managing risk at the firm level as well as ways to improve productivity and performance.

The Center fosters the development of a community of faculty, visiting scholars and Ph.D. candidates whose research interests complement and support the mission of the Center. The Center works closely with industry executives and practitioners to ensure that its research is informed by the operating realities and competitive demands facing industry participants as they pursue competitive excellence.

Copies of the working papers summarized here are available from the Center. If you would like to learn more about the Center or become a member of our research community, please let us know of your interest.

Anthony M. Santomero
Director

The Working Paper Series is made possible by a generous grant from the Alfred P. Sloan Foundation

Inside the Black Box:
What Explains Differences in the Efficiencies of Financial Institutions? [1]

January 8, 1997

Forthcoming in *Journal of Banking and Finance*

Abstract: Over the past several years, substantial research effort has gone into measuring the efficiency of financial institutions. Many studies have found that inefficiencies are quite large, on the order of 20% or more of total banking industry costs and about half of the industry's potential profits. There is no consensus on the sources of the differences in measured efficiency. This paper examines several possible sources, including differences in efficiency concept, measurement method, and a number of bank, market, and regulatory characteristics. We review the existing literature and provide new evidence using data on U.S. banks over the period 1990-95.

JEL Classification Numbers: G2, D2, G21, G28, E58, E61, F33

Keywords: Bank, efficiency, cost, profit

---

[1]Allen N. Berger is at the Board of Governors of the Federal Reserve System, Washington, DC 20551 U.S.A., and Wharton Financial Institutions Center, Philadelphia, PA 19104 U.S.A.

Loretta J. Mester is at the Federal Reserve Bank of Philadelphia, Philadelphia, PA 19106 U.S.A., and the Finance Department, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104 U.S.A.

The views expressed in this paper do not necessarily represent those of the Federal Reserve Bank of Philadelphia, of the Board of Governors of the Federal Reserve System, or of the Federal Reserve System. The authors thank Emilia Bonaccorsi, Dave Humphrey and audience participants at the Federal Reserve Bank of New York and the Atlantic Economic Society meetings for helpful comments, and Seth Bonime, Margaret Kyle, and Joe Scalise for excellent research assistance.

Correspondence to Berger at Mail Stop 153, Federal Reserve Board, 20th and C Sts. N.W., Washington, D.C. 20551; phone: (202)-452-2903; fax: (202) 452-5295; email: m1anb00@frb.gov. To Mester at Research Department, Federal Reserve Bank of Philadelphia, Ten Independence Mall, Philadelphia, PA 19106-1574; phone: (21 5) 574-3807; fax: (215) 574-4364; email: lmester@frbphil.org.

Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?

1. Introduction

Over the past several years, substantial research effort has gone into measuring the efficiency of financial institutions, particularly commercial banks. The focus has been on estimating an efficient frontier and measuring the average differences between observed banks and banks on the frontier. Many studies have found large inefficiencies, on the order of 20% or more of total banking industry costs, and about half of the industry's potential profits. There is no consensus on the sources of the differences in measured efficiency. An obvious next step in the efficiency research program is to determine these sources. This paper focuses on three sources: (1) differences in the efficiency concept used; (2) differences in measurement methods used to estimate efficiency within the context of these concepts; and (3) potential correlates of efficiency—bank, market, and regulatory characteristics that are at least partially exogenous and may explain some of the efficiency differences that remain after controlling for efficiency concept and measurement method. We review the existing literature on the sources of efficiency of financial institutions and provide new evidence.

Estimates of efficiency often vary substantially across studies according to the data source, as well as the efficiency concepts and measurement methods used in the studies. Berger and Humphrey (1997) documented 130 studies of financial institution efficiency, using data from 21 countries, from multiple time periods, and from various types of institutions including banks, bank branches, savings and loans, credit unions, and insurance companies. These variations in the data sets from which efficiencies are measured make it virtually impossible to determine how important the different efficiency concepts, measurement techniques, and correlates used are to the outcomes of these studies. Put another way, the sources of differences in efficiency across financial institutions are concealed from view within an opaque "black box" because the individual studies simultaneously differ from one another in so many different dimensions.

Our empirical application tries to get around this problem by employing multiple efficiency concepts, using a number of different measurement methods, and applying a comprehensive set of potential efficiency correlates to a single data set. We estimate the efficiency of almost 6,000 U.S. commercial banks that were in continuous existence over the six-year period 1990-95 and had no missing or questionable data on any of the variables used. Thus, the differences we observe should reasonably accurately reflect the effects of changes in the concepts, measurement techniques, and potential correlates that are used, rather than any differences in the data set to which these assumptions are applied.

We employ three distinct economic efficiency concepts—cost, standard profit, and alternative profit efficiencies. We analyze the effects of a number of measurement methods, including use of the distribution-free approach versus the stochastic frontier approach, specification of the Fourier-flexible functional form versus the translog form, and inclusion of problem loans and financial capital in a number of different ways. We find that measured efficiency differs across the three efficiency concepts, and that each adds some independent informational value. A somewhat surprising result is that the choices made concerning efficiency measurement usually make very little difference to our empirical findings in terms of either average industry efficiency or rankings of individual firms, suggesting that the efficiency estimates are fairly robust to differences in methodology. Another surprising result is that we also find substantial unexploited cost scale economies for fairly large sizes of banks in the 1990s, suggesting a change from the 1980s.

Once the conceptual and measurement issues have been controlled for, it is important for the purposes of public policy, research, and managerial performance to explain the remaining differences in efficiency across banks. In a perfectly competitive or contestable market, one would expect inefficient firms to be driven out by efficient firms, so that there would be only a residual level of inefficiency across firms remaining at any given time. An empirical finding of substantial inefficiencies, therefore, raises the question as to whether inefficiencies, which may have been sustainable in the past because of regulatory limits on competition, will continue in the less-regulated future. For antitrust and merger analysis, it is important to know the effects of market concentration and past mergers on banking efficiency. Similarly, it is important to know whether one type of organizational form is more efficient than another, and whether inefficiency manifests itself in the form of poor production decisions, risk management decisions, or both. We review the existing studies that analyzed potential correlates of efficiency, but a comparison across studies is hampered by the fact that different samples, efficiency concepts, and measurement techniques were used. In our empirical analysis, we explore the effects of a number of potential correlates of bank efficiency after controlling for efficiency concept and measurement method. The potential correlates include measures of bank size, organizational form and corporate governance, other bank characteristics, market characteristics, state geographic restrictions, and federal regulator. We find that a number of these factors appear to have independent influences on efficiency, although many expected effects are not present and some of the effects we find are not consistent with expectations.

2. The Efficiency Concept—Cost, Standard Profit, and Alternative Profit Efficiency

A fundamental decision in measuring financial institution efficiency is which concept to use. This, of course, depends on the question being addressed. We discuss here what we consider to be the three most important economic efficiency concepts—cost, standard profit, and alternative profit efficiencies. We believe these concepts have the best economic foundation for analyzing the efficiency of financial institutions because they are based on economic optimization in reaction to market prices and competition, rather than being based solely on the use of technology.

2.1 Cost Efficiency. Cost efficiency gives a measure of how close a bank's cost is to what a best-practice bank's cost would be for producing the same output bundle under the same conditions. It is derived from a cost function in which variable costs depend on the prices of variable inputs, the quantities of variable outputs and any fixed inputs or outputs, environmental factors, and random error, as well as efficiency. Such a cost function may be written as:

$$C = C(w,y,z,v,u_C,\epsilon_C), \qquad (1)$$

where C measures variable costs, w is the vector of prices of variable inputs, y is the vector of quantities of variable outputs, z indicates the quantities of any fixed netputs (inputs or outputs), which are included to account for the effects of these netputs on variable costs owing to substitutability or complementarily with variable netputs, v is a set of environmental or market variables that may affect performance, u_C denotes an inefficiency factor that may raise costs above the best-practice level, and ϵ_C denotes the random error that incorporates measurement error and luck that may temporarily give banks high or low costs. The inefficiency factor u_C incorporates both allocative inefficiencies from failing to react optimally to relative prices of inputs, w, and technical inefficiencies from employing too much of the inputs to produce y. To simplify the measurement of efficiency, the inefficiency and random terms u_C and ϵ_C are assumed to be multiplicatively separable from the rest of the cost function, and both sides of (1) are represented in natural logs:

$$ln\, C = f(w,y,z,v) + ln\, u_C + ln\, \epsilon_C, \tag{2}$$

where f denotes some functional form. The term, $ln\, u_C + ln\, \epsilon_C$, is treated as a composite error term, and the various X-efficiency measurement techniques (described in section 3.1) differ in how they distinguish the inefficiency term, $ln\, u_C$, from the random error term, $ln\, \epsilon_C$. We define the cost efficiency of bank b as the estimated cost needed to produce bank b's output vector if the bank were as efficient as the best-practice bank in the sample facing the same exogenous variables (w,y,z,v) divided by the actual cost of bank b, adjusted for random error, i.e.,

$$\text{Cost EFF}^{b} = \frac{\hat{C}^{min}}{\hat{C}^{b}} = \frac{\exp[\hat{f}(w^b,y^b,z^b,v^b)] \times \exp[ln\, \hat{u}_C^{min}]}{\exp[\hat{f}(w^b,y^b,z^b,v^b)] \times \exp[ln\, \hat{u}_C^{b}]} = \frac{\hat{u}_C^{min}}{\hat{u}_C^{b}}, \tag{3}$$

where $\hat{u}_C^{min}$ is the minimum $\hat{u}_C^{b}$ across all banks in the sample.

The cost efficiency ratio may be thought of as the proportion of costs or resources that are used efficiently. For example, a bank with Cost EFF of 0.70 is 70% efficient or equivalently wastes 30% of its costs relative to a best-practice firm facing the same conditions. Cost efficiency ranges over (O, 1], and equals one for a best-practice firm within the observed data.[1]

2.2 Standard Profit Efficiency. Standard profit efficiency measures how close a bank is to producing the maximum possible profit given a particular level of input prices and output prices (and other

[1]In applications, efficiency is generally defined relative to the best practice observed in the industry, rather than to any true minimum costs, since the underlying technology is unknown. (The usual form of the stochastic frontier measurement technique is an exception.) Fortunately, for most economic hypotheses, relative efficiency rather than absolute efficiency is the more appropriate concept. For example, we investigate below whether larger versus smaller banks are more efficient, which requires only comparisons to a consistent frontier.

variables). In contrast to the cost function, the standard profit function specifies variable profits in place of variable costs and takes variable output prices as given, rather than holding all output quantities statistically fixed at their observed, possibly inefficient, levels. That is, the profit dependent variable allows for consideration of revenues that can be earned by varying outputs as well as inputs. Output prices are taken as exogenous, allowing for inefficiencies in the choice of outputs when responding to these prices or to any other arguments of the profit function.

The standard profit function, in log form, is:

$$ln(\pi + \theta) = f(w,p,z,v) + ln\, u_\pi + ln\, \epsilon_\pi , \tag{4}$$

where π is the variable profits of the firm, which includes all the interest and fee income earned on the variable outputs minus variable costs, C, used in the cost function; θ is a constant added to every firm's profit so that the natural log is taken of a positive number; p is the vector of prices of the variable outputs; $ln\, \epsilon_\pi$ represents random error; and $ln\, u_\pi$ represents inefficiency that reduces profits.

We define standard profit efficiency as the ratio of the predicted actual profits to the predicted maximum profits that could be earned if the bank was as efficient as the best bank in the sample, net of random error, or the proportion of maximum profits that are actually earned:

$$\text{Std } \pi \text{ EFF}^b = \frac{\hat{\pi}^b}{\hat{\pi}^{max}} = \frac{\left\{\exp[\hat{f}(w^b,p^b,z^b,v^b)] \times \exp[ln\, \hat{u}_\pi^b]\right\} - \theta}{\left\{\exp[\hat{f}(w^b,p^b,z^b,v^b)] \times \exp[ln\, \hat{u}_\pi^{max}]\right\} - \theta} \tag{5}$$

where $\hat{u}_\pi^{max}$ is the maximum value of u_π^b in the sample.[2]

Standard profit efficiency is the proportion of maximum profits that are earned, so that a Std π EFF ratio of 0.70 would indicate that, because of excessive costs, deficient revenues, or both, the firm is losing about 30% of the profits it could be earning. Similar to the cost efficiency ratio. the profit efficiency ratio

[2]The profit efficiency does not simplify to a ratio of $\hat{u}_\pi$'s as in the case of cost efficiency because the addition of θ to the dependent variable before taking logs means that the efficiency factor is not exactly multiplicatively separable in the profit function. A bank's efficiency will vary somewhat with the values of the exogenous variables, so for our efficiency estimates we average the values of the numerator and denominator in (5) over the sample period before dividing to measure the average efficiency of the bank over the sample period.

equals one for a best-practice firm that maximizes profits for its given conditions within the observed data. Unlike cost efficiency, however, profit efficiency can be negative, since firms can throw away more than 100% of their potential profits.

In our opinion, the profit efficiency concept is superior to the cost efficiency concept for evaluating the overall performance of the firm. Profit efficiency accounts for errors on the output side as well as those on the input side, and some prior evidence suggested that inefficiencies on the output side may be as large or larger than those on the input side (e.g., Berger, Hancock, and Humphrey, 1993). Profit efficiency is based on the more accepted economic goal of profit maximization, which requires that the same amount of managerial attention be paid to raising a marginal dollar of revenue as to reducing a marginal dollar of costs. That is, a firm that spends $1 additional to raise revenues by $2, all else held equal, would appropriately be measured as being more profit efficient but might inappropriately be measured as being less cost efficient.

Profit efficiency is based on a comparison with the best-practice point of profit maximization within the data set, whereas cost efficiency evaluates performance holding output constant at its current level, which generally will not correspond to an optimum. A firm that is relatively cost efficient at its current output may or may not be cost efficient at its optimal output, which typically involves a different scale and mix of outputs. Thus, standard profit efficiency may take better account of cost inefficiency than the cost efficiency measure itself, since standard profit efficiency embodies the cost inefficiency deviations from the optimal point.[3]

2.3 Alternative Profit Efficiency. An interesting recent development in efficiency analysis is the concept of alternative profit efficiency, which may be helpful when some of the assumptions underlying cost and standard profit efficiency are not met. Efficiency here is measured by how close a bank comes to earning

[3]A few prior papers have studied standard profit efficiency at U.S. banks (Berger, Hancock, and Humphrey, 1993, DeYoung and None, 1996, Akhavein, Swamy, and Taubman, 1997, and Akhavein, Berger, and Humphrey, 1997). The measured average profit efficiencies ranged from 24% of potential profits being earned to 67%. Profit function estimation was also used to measure efficiency in terms of the risk-expected return efficient frontier as defined in the finance literature (Hughes and Moon, 1995, Hughes, Lang, Mester, and Moon, 1996a,b). A bank with too little expected profit for the amount of risk it is taking on is deemed inefficient. Average efficiency in terms of the percent of expected profit being earned for a given level of risk relative to the best practice banks was found to be around 85%.

maximum profits given its output levels rather than its output prices. The alternative profit function employs the same dependent variable as the standard profit function and the same exogenous variables as the cost function. Thus, instead of counting deviations from optimal output as inefficiency, as in the standard profit function, variable output is held constant as in the cost function while output prices are free to vary and affect profits. The alternative profit function in log form is:

$$ln(\pi + \theta) = f(w,y,z,v) + ln\, u_{a\pi} + ln\, \epsilon_{a\pi}, \tag{6}$$

which is identical to the standard profit function in (3) except that y replaces p in the function, f, yielding different values for the inefficiency and random error terms, $ln\, u_{a\pi}$ and $ln\, \epsilon_{a\pi}$, respectively.

As with standard profit efficiency, alternative profit efficiency is the ratio of predicted actual profits to the predicted maximum profits for a best-practice bank:

$$\text{Alt } \pi \text{ EFF}^{b} = \frac{a\hat{\pi}^{b}}{a\hat{\pi}^{max}} = \frac{\left\{\exp[\hat{f}(w^{b},y^{b},z^{b},v^{b})] \times \exp[ln\, \hat{u}_{a\pi}^{b}]\right\} - \theta}{\left\{\exp[\hat{f}(w^{b},y^{b},z^{b},v^{b}) \times \exp[ln\, \hat{u}_{a\pi}^{max}]\right\} - \theta}\ . \tag{7}$$

Here, efficiency values are allowed to vary in an important way with output prices, but errors in choosing output quantities do not affect alternative profit efficiency except through the point of evaluation $\hat{f}(w^{b},y^{b},z^{b},v^{b})$ to the extent that the best-practice bank is not operating at the same (w,y,z,v) as bank b.

There would be no reason to estimate alternative profit efficiency if the usual assumptions held. Standard profit efficiency and cost efficiency would appropriately measure how well the firm was producing outputs and employing inputs relative to best-practice firms, given the underlying assumptions. However, alternative profit efficiency may provide useful information when one or more of the following conditions hold:

(i) there are substantial unmeasured differences in the quality of banking services;

(ii) outputs are not completely variable, so that a bank cannot achieve every output scale and product mix;

(iii) output markets are not perfectly competitive, so that banks have some market power over the prices they charge; and

(iv) output prices are not accurately measured, so they do not provide accurate guides to opportunities to earn revenues and profits in the standard profit function.

The alternative profit function provides a way of controlling for unmeasured differences in output quality, as in condition (i), since it considers the additional revenue that higher quality output can generate. If output markets are competitive and customers are willing to pay for the additional services provided by some banks, these banks should receive higher revenues that just compensate for their extra costs. Banks would be sorted into market niches that differ by service quality or intensity, with customers who need or prefer higher quality or more service paying more per dollar of their loan or deposit. Since the higher interest rates or fees received by the higher quality providers just cover their extra production costs, these banks survive in competitive equilibrium. For example. banks that take on more information-problematic loans should charge higher interest rates or fees to cover their extra origination, monitoring, and control costs than banks that lend to equally risky, but more informationally transparent borrowers. The alternative profit function essentially replicates the cost function except that it adds revenues to the dependent variable. It accounts for the additional revenue earned by high-quality banks, allowing it to offset their additional costs of providing the higher service levels. So it does not penalize high-quality banks in terms of their efficiency measure, whereas the cost function might. Thus, if banks do not have market power, alternative profit efficiency should be thought of as a better measure of cost efficiency, rather than profit efficiency, since it does not take into account any errors in the quantities of variable outputs.[4] Other methods of controlling for differences in output quality are discussed in section 3.3 below.

Alternative profit efficiency might also prove useful if the variable outputs are not completely variable, as in condition (ii) above. Banks differ in size by more than 1000-fold, even within the same local markets. Most banks have fewer than $100 million in assets, yet they operate side-by-side with megabanks with over $100 billion in assets. Clearly, a bank below $100 million cannot reach the size of a megabank except after

---

[4]Differences in output quality may also be partially captured in the standard profit function. However, since it holds output prices fixed, the standard profit function is less able to account for differences in revenue that compensate for differences in product quality, since these revenue differences may be partly reflected in measured prices. Berger, Cummins, and Weiss (1996) found that both standard and alternative profit efficiencies helped control for differences in service quality in property-liability insurance industry.

decades of growth and mergers and acquisitions, yet the standard profit function essentially treats these large and small banks as if they should have the same variable outputs when facing the same input and output prices, fixed netputs, and environmental variables specified in the standard profit function. Thus, unless the (w,p,z,v) variables give a strong prediction about the size of the bank, a scale bias may occur in the standard profit function, as larger banks have higher profits that are not explained by the exogenous variables. That is, large banks may (arguably mistakenly) be labeled as having higher standard profit efficiency than smaller banks, by virtue of the fact that small banks simply cannot reach the same output levels. This potential problem does not occur to the same degree for the alternative profit function, since outputs are held constant statistically. That is, alternative profit efficiency compares the ability of banks to generate profits for the same levels of output and therefore reduces the scale bias that might be present in the standard profit efficiency measure.

The alternative profit efficiency concept may also be helpful in situations in which the firms exercise some market power in setting output prices, as in condition (iii). The standard profit function takes output prices as given and embodies the assumption that the bank can sell as much output as it wishes without having to lower its prices. This can lead to an understatement of standard profit efficiency for firms with output below efficient scale, since these firms might have to reduce their prices to increase output and, therefore, cannot earn as much as maximum potential profits as we measure it.[5]

Under conditions of market power, it may be appropriate to consider output levels as relatively fixed in the short run and allow for efficiency differences in the setting of prices and service quality. That is, an optimizing bank will set each of its prices at the point where the market just clears for its output and choice of service quality. Such a bank will also choose an optimizing service quality niche. Unlike the perfect

---

[5]Empirical studies have shown that banks with larger shares of the local market have some control over prices, paying lower rates to small depositors (Berger and Hannan, 1989) and charging higher rates to small borrowers (Hannan, 1991). These results are supported by studies that have tested price-taking versus price-setting behavior for banks, most often finding the latter (Hancock, 1986, Hannan and Liang, 1993, and English and Hayes, 1991). Berger, Humphrey, and Pulley (1996) estimated that about 68% of U.S. bank revenues are from products competed for on a local basis and, therefore, could be subject to price-setting behavior. However, it is not known how many of the prices of these products actually do contain significant market power premiums.

competition case considered above, a firm with market power may be able to increase revenues more than costs by increasing service quality because there may not be other competitors or potential competitors at that quality niche. It is also possible that the optimizing choice may be to economize on service quality and keep costs relatively low. Alternative profit efficiency measures the extent to which firms are able to optimize in their choices of prices and service quality, as well as their abilities to keep costs low for a given output level. Alternative profit efficiency will also incorporate differences across firms in market power and their abilities to exploit it, which is good for the owners of the bank, but is not a social good in the same way that the other efficiencies are. Alternative profit efficiency may be viewed as a robustness check on standard profit efficiency, which takes prices as fixed and allows outputs to be totally variable.

The measurement of alternative profit efficiency may also be motivated in part by inaccuracies in the output price data, as in condition (iv) above. If the output price vector, p, is well measured, it should be strongly related to profits and explain a substantial portion of the variance of profits in the standard profit function. If prices are inaccurately measured—as is likely given the available banking data—the predicted part of the standard profit function, f, in (4) would explain less of the variance of profits and yield more error in the estimation of the efficiency term $ln\ u_{\pi}$.[6] In this event, it may be appropriate to try specifying other variables in the profit function that might yield a better fit, such as the output quantity vector, y, as in the alternative profit function.[7]

---

[6]There are good reasons to believe that output prices may be inaccurately measured in banking data. Regulatory reports, such as the Call Report form, require accurate figures on balance-sheet quantities, but do not directly measure prices. Rather, prices used in efficiency studies often must be constructed as ratios of revenue flows to stocks of assets, which may incorporate noise due to differences in asset duration, risk, liquidity, collateral, etc., as well as problems in matching revenue flows with the assets and time periods on which they were earned.

[7]One way to examine the problem of inaccurate price data is to determine the extent to which measured prices help predict profits in the profit function. Humphrey and Pulley (1997) specified a bank profit function with both prices, p, and quantities, y, included. A test of the joint hypothesis that all the p parameters were zero was not rejected by the data, whereas the data did reject the hypothesis that all the y parameters were zero. These results suggest that measured output prices do **not** have the theoretically predicted strong positive relationship with profits, and that output quantities do strongly predict profits, perhaps in part reflecting the scale bias problem discussed above that output quantities are not completely variable over the short term.

Another possible specification of the profit function would be to include **neither** output prices, p, nor

3. Efficiency Measurement Methods

Once the efficiency concepts are selected, the next issue is how to go about measuring them. Here we explore four methodological choices—the estimation technique, the functional form specified (assuming a parametric technique is chosen), the treatment of output quality, and the role of financial capital.

3.1 Estimation Techniques. The most common efficiency estimation techniques are data envelopment analysis (DEA), free disposable hull analysis (FDH), the stochastic frontier approach, the thick frontier approach, and the distribution-free approach.[8] The first two of these are nonparametric techniques and the latter three are parametric methods. Berger and Humphrey (1997) reported roughly an equal split between applications of nonparametric techniques (69 applications) and parametric methods (60 applications) to depository institutions data.

Here, we focus on the parametric techniques primarily because they correspond well with the cost and profit efficiency concepts outlined above. The nonparametric methods generally ignore prices and can, therefore, account only for technical inefficiency in using too many inputs or producing too few outputs. They cannot account for allocative inefficiency in misresponding to relative prices in choosing inputs and outputs, nor can they compare firms that tend to specialize in different inputs or outputs, because there is no way to compare one input or output with another without the benefit of relative prices. In addition, similar to the cost function, there is no way to determine whether the output being produced is optimal without value information on the outputs. Thus, the nonparametric techniques typically focus on **technological** optimization rather than **economic** optimization, and do not correspond to the cost and profit efficiency concepts discussed above. Another drawback of the nonparametric techniques is that they usually do not allow for random error in the data, assuming away measurement error and luck as factors affecting outcomes (although some progress is

---

quantities, y. Efficiency would be measured relative to a frontier in which firms optimize over output prices, quantities, and service quality jointly. As argued by Berger, Humphrey, and Pulley (1996) and Humphrey and Pulley (1997), such a specification would likely be too sparse to describe the conditions faced by individual banks and would also be subject to scale biases. It is essentially rejected by the data in Humphrey and Pulley's (1997) test of the y parameters in the profit function.

[8] See Mester (1994) for further description of these techniques.

being made in this regard). In effect, they disentangle efficiency differences from random error by assuming that random error is zero. Studies of U.S. banks that use nonparametric techniques report lower efficiency means on average than those using parametric techniques (an average of 72% versus 84%) with much greater variation (a standard deviation of 17% versus 6%), which could, in part, reflect some random error being counted as variations in measured efficiency in these studies (Berger and Humphrey 1997, Table 2).

In the parametric methods, a bank is labeled inefficient if its costs are higher or profits are lower than the best-practice bank after removing random error—in other words, if the estimated $ln\ u_C$, $ln\ u_\pi$, $ln\ u_{a\pi}$, in equations (2), (3), and (4), respectively, differ substantially from the best-practice values.[9] The methods differ in the way $ln\ u$ is disentangled from the composite error term $ln\ u + ln\ \epsilon$. In our study we use both the stochastic frontier approach and the distribution-free approach. As discussed below, the distribution-free approach is our preferred technique.

In the **stochastic frontier approach,** the inefficiency and random error components of the composite error term are disentangled by making explicit assumptions about their distributions. The random error term, $ln\ \epsilon$, is assumed to be two-sided (usually normally distributed), and the inefficiency term, $ln\ u$, is assumed to be one-sided (usually half-normally distributed). The parameters of the two distributions are estimated and can be used to obtain estimates of bank-specific inefficiency. The estimated mean of the conditional distribution of $ln\ u$ given $ln\ u + ln\ \epsilon$, i.e., $ln\ \hat{u} \equiv \hat{E}(ln\ u | ln\ u + ln\ \epsilon)$ is usually used to measure inefficiency.

The distributional assumptions of the stochastic frontier approach are fairly arbitrary. Two prior studies found that when the inefficiencies were unconstrained, they behaved much more like symmetric normal distributions than half-normals, which would invalidate the identification of the inefficiencies (Bauer and Hancock 1993, Berger 1993).[10] As shown below, the data in the current study are often consistent with the

[9]In the typical application of the stochastic frontier approach, inefficiency is measured relative to the estimated frontier, f, rather than the best-practice bank, i e., relative to a zero value for $ln\ u$, which is not achieved by any firm in the sample. To make our efficiency measures comparable across techniques, we normalize our stochastic frontier efficiency estimates to be deviations from the smallest observed expected value of $ln\ u$, so that the most efficient bank in the sample has efficiency of one.

[10]Other distributions have also been used, e.g., normal–truncated normal (Stevenson, 1980, Mester, 1996, Berger and DeYoung, 1996), normal–gamma (Stevenson, 1980, and Greene, 1990), and

presence of this potential problem—in many cases, the residuals are simply not skewed in the direction predicted by the assumptions of the stochastic frontier approach.

If panel data are available, some of these maintained distributional assumptions can be relaxed, and the **distribution-free approach** may be used. This method assumes that there is a core efficiency or average efficiency for each firm over time. The core inefficiency is distinguished from random error (including any temporary fluctuations in efficiency) by assuming that core inefficiency is persistent over time, while random errors tend to average out over time. In particular, a cost or profit function is estimated for each period of a panel data set. The residual in each separate regression is composed of both inefficiency, $\ln u$, and random error, $\ln \epsilon$, but the random component, $\ln \epsilon$, is assumed to average out over time, so that the average of a bank's residuals from all of the regressions, $\ln \hat{u}$, will be an estimate of the inefficiency term, $\ln u$. For banks with very low or very high $\ln \hat{u}$, an adjustment (called truncation) is made to assign less extreme values of $\ln \hat{u}$ to these banks, since extreme values may indicate that random error, $\ln \epsilon$, has not been completely purged by averaging. The resulting $\ln \hat{u}$ for each bank is used to compute its core efficiency.[11]

3.2 Functional Forms for the Parametric Methods. We next consider the choice of a functional form for the cost and profit functions, f, when one of the parametric methods is used to estimate efficiency. The most popular form in the literature is the translog; however, it does not necessarily very well fit data that are far from the mean in terms of output size or mix. McAllister and McManus (1993), and Mitchell and Onvural (1996) showed that some of the differences in results on scale economies across studies may be due to the ill-fit of the translog function across a wide range of bank sizes, some of which may be underrepresented in the data.

normal–exponential (Mester, 1996).

[11]The reasonableness of these assumptions about the error term components depends on the length of period studied. If too short a period is chosen, the random errors might not average out, in which case random error would be attributed to inefficiency (although truncation can help). If too long a period is chosen, the firm's core efficiency becomes less meaningful because of changes in management and other events, i.e., it might not be constant over the time period. Using 1984-94 data on U.S. commercial banks and assuming a translog cost model, DeYoung (1997) showed that a six-year time period, such as we use here, reasonably balanced these concerns.

A more flexible functional form would help to alleviate this problem. The Fourier-flexible functional form augments the translog by including Fourier trigonometric terms. It is more flexible than the translog and is a global approximation to virtually any cost or profit function. Several studies have shown that it fits the data for U.S. financial institutions better then the translog.[12] Berger and DeYoung (1996) found that measured inefficiencies were about twice as large when the translog was specified in place of the Fourier-flexible form.[13] Here, we estimate the Fourier-flexible functional form and allow our cost and profit frontiers to vary each year, but also evaluate the effects of switching to the translog by restricting the Fourier terms to be zero.[14]

3.3 Output Quality. Theoretically, in comparing one bank's efficiency to another's, the comparison should be between banks producing the **same** output quality. But there are likely to be unmeasured differences in quality because the banking data does not fully capture the heterogeneity in bank output. The amount of service **flow** associated with financial products is by necessity usually assumed to be proportionate to the dollar value of the **stock** of assets or liabilities on the balance sheet, which can result in significant mismeasurement. For example, commercial loans can vary in size, repayment schedule, risk, transparency of information, type of collateral, covenants to be enforced, etc. These differences are likely to affect the costs to the bank of loan origination, ongoing monitoring and control, and financing expense. Unmeasured differences in product

[12] See McAllister and McManus (1993), Berger, Cummins, and Weiss (1996), Berger and DeYoung (1996), Berger, Leusner, and Mingo (1996), and Mitchell and Onvural (1996). McAllister and McManus (1993) also used kernel regression and spline estimation techniques to obtain better global properties.

[13] Other functional forms have also been specified. Mester (1992) estimated a hybrid translog function, and Berger, Hancock, and Humphrey (1993) estimated a Fuss normalized quadratic variable profit function. Hughes, Lang, Mester, and Moon (1995, 1996a,b) estimated a utility-maximization model based on the Almost Ideal Demand System consisting of profit and input share equations. If risk neutrality is imposed on this system, it corresponds to the standard translog cost function and input share equations.

[14] To further increase flexibility, one can allow the parameters being estimated to differ across banks that may be using different production technologies, e.g., banks of different sizes, banks facing different regulatory regimes, banks operating in different time periods, or different types of institutions. Numerous studies have allowed the coefficients to vary according to whether the bank operates in a state that restricts branching or a state that allows intrastate branching (e.g., Berger, 1993). Mester (1993) found a significant difference in both the frontier parameters and parameters of the error term distribution in the stochastic frontier method for mutual and stock-owned savings and loans. Most studies using the distribution-free method allow the frontier parameters to vary over time. Akhavein, Swamy, and Taubman (1997) used random coefficient estimation techniques, which allow each bank to have its own parameters.

quality may be incorrectly measured as differences in cost inefficiency.

We have already discussed how the alternative profit function can help control for unmeasured differences in output quality. Other studies took another approach and included variables intended to control for the quality of bank output. For example, Hughes and Mester (1993), Hughes, Lang, Mester, and Moon (1996a,b) and Mester (1996) included the volume of nonperforming loans as a control for loan quality in studies of U.S. banks, and Berg, Førsund, and Jansen (1992) included loan losses as an indicator of the quality of loan evaluations in a DEA study of Norwegian bank productivity.

Whether it is appropriate econometrically to include nonperforming loans and loan losses in the bank's cost, standard profit, and alternative profit functions depends on the extent to which these variables are exogenous. Nonperforming loans and loan losses would be exogenous if caused by negative economic shocks ("bad luck"), but they could be endogenous, either because management is inefficient in managing its portfolio ("bad management") or because it has made a conscious decision to reduce short-run expenses by cutting back on loan origination and monitoring resources ("skimping").[15] Berger and DeYoung (1996) tested the "bad luck," "bad management," and "skimping" hypotheses and found mixed evidence on the exogeneity of nonperforming loans. In our empirical analysis below we attempt to solve this problem using the ratio of nonperforming loans to total loans in the bank's state. Our state average variable is almost entirely exogenous to any individual bank, but allows us to control for negative shocks that may affect the bank.

3.4 The Role of Financial Capital. Another aspect of efficiency measurement is the treatment of financial capital. A bank's insolvency risk depends on its financial capital available to absorb portfolio losses, as well as on the portfolio risks themselves. Insolvency risk affects bank costs and profits via the risk premium the bank has to pay for uninsured debt, and through the intensity of risk management activities the bank undertakes. For this reason, the financial capital of the bank should be considered when studying

---

[15]Of course, even if the level of nonperforming loans does reflect bank choice to some extent, it could still be appropriate to include it in the cost and profit functions if it is thought to reflect a less frequent decision on the part of the bank (e.g., credit policy) than production decisions. This is the same logic that allows the output levels, which are ultimately endogenous variables chosen by the bank, to be included in the cost and alternative profit functions.

efficiency. To some extent, controlling for the interest rates paid on uninsured debt helps account for differences in risk, but these rates are imperfectly measured.

Even apart from risk, a bank's capital level directly affects costs by providing an alternative to deposits as a finding source for loans. Interest paid on debt counts as a cost, but dividends paid do not. On the other hand, raising equity typically involves higher costs than raising deposits. If the first effect dominates, measured costs will be higher for banks using a higher proportion of debt financing; if the second effect dominates, measured costs will be lower for these banks. Large banks depend more on debt financing to finance their portfolios than small banks do, so a failure to control for equity could yield a scale bias.

The specification of capital in the cost and profit functions also goes part of the way toward accounting for different risk preferences on the part of banks. The cost, standard profit, and alternative profit efficiency concepts discussed in section 2 take as given that banks are risk neutral. But if some banks are more risk averse than others, they may hold a higher level of financial capital than maximizes profits or minimizes costs. If financial capital is ignored, the efficiency of these banks would be mismeasured, even though they are behaving optimally given their risk preferences. Hughes, Lang, Mester, and Moon (1995, 1996a,b) and Hughes and Moon (1995) tested and rejected the assumption of risk neutrality for banks.

Despite these arguments, only a few efficiency studies have included financial capital. Hancock (1985, 1986) conditioned an average-practice profit function on financial capital. Clark (1996) included capital in a model of economic cost and found that it eliminated measured scale diseconomies in production costs alone. The Hughes and Mester (1993, 1996) cost studies and the Hughes, Lang, Mester, and Moon (1995, 1996a) profit studies incorporated financial capital and found increasing returns to scale at large-asset-size banks, unlike studies that did not incorporate capital. One possible reason is that large size confers diversification benefits that allow large banks to have lower capital ratios than smaller banks. Akhavein, Berger, and Humphrey (1997) controlled for equity capital and found that profit efficiency increases as a result of mergers of large banks. Merged banks tend to shift their portfolios toward loans and away from securities for a given level of equity. This could reflect diversification benefits available to merged banks—better diversification would allow the merged bank to manage better the increased portfolio risk with the same amount of equity capital. In the efficiency estimates presented below, we incorporate financial capital in the cost and profit function specifications.

4. Efficiency Correlates

Once we have controlled for the efficiency concepts and measurement methods used, and applied these concepts and methods to the same data set, what explains the remaining differences in efficiency across banks? The answer to this question has important implications for public policy, research, and bank management. A useful first step is to explore the effects of a number of potential correlates of bank efficiency—various bank, market, and regulatory characteristics that are at least partially exogenous to efficiency and so may help explain the observed large differences in efficiency across banks. Several papers have performed analyses along these lines.[16] A two-step procedure is typically used, whereby firm efficiency is estimated using one of the techniques described above and is then regressed on, or tested for correlation with, a set of variables describing the characteristics being investigated.[17]

Some econometric issues make such analyses suggestive but not conclusive. First, the dependent variable in the regressions, efficiency, is an estimate, but the standard error of this estimate is not accounted for in the subsequent regression or correlation analysis. Second, none of the variables used in the regressions is completely exogenous, and the endogeneity of any regressor can bias the coefficient estimates on all the regressors. Even a characteristic like the identity of the bank's primary federal regulator is somewhat endogenous, since banks can change their charters. Endogeneity makes conclusions about causation problematic. As an alternative to regression analysis, simple correlations are provided in some papers to underscore the fact that causation may run in both directions.

---

[16]Bank studies include Aly, Grabowski, Pasurka, and Rangan (1990), Berger, Hancock, and Humphrey (1993), Pi and Timme (1993), Kaparakis, Miller, and Noulas (1994), Berger and Hannan (1996), Kwan and Eisenbeis (1995), Spong, Sullivan, and DeYoung (1995), Hughes, Lang, Mester, and Moon (1996a,b), and Mester (1996); savings and loan studies include Cebenoyan, Cooperman, Register, and Hudgins (1993), Mester (1993), and Hermalin and Wallace (1994).

[17]The regressions are usually linear, but Mester (1993, 1996) used the logistic functional form, as the stochastic frontier inefficiency estimates varied between zero and one.

The different measurement techniques and efficiency concepts used and time periods and samples studied make it difficult to compare the results of the regression analyses across studies. The potential correlates used in the second-stage regressions also vary substantially across studies, sometimes because each study has a particular focus—e. g., market structure, geographic diversification, or corporate control.

Most studies included the asset size of the institution, but no consistent picture emerges of its relationship with efficiency.[18] Evidence on organizational form was also mixed.[19] There is weak evidence that banks in holding companies are more efficient than independent banks.[20] The relationship between the size of the CEO's stock ownership and efficiency varies across studies.[21] There is limited evidence that banks operating in more concentrated markets are less efficient, supporting the "quiet life" theory that inefficiency has been sustainable in banking because competition has not been robust.[22]

Most of the studies have found that well-capitalized banks and S&Ls are more efficient. This is consistent with moral hazard theory that suggests managers of institutions closer to bankruptcy might be inclined to pursue their own interests. But causation could run the other way—less efficient institutions have lower profits, leading to lower capital ratios. Another fairly general finding among the bank studies is that more efficient banks have lower levels of nonperforming loans, but as described above, nonperforming loans likely have exogenous and endogenous components.[23] As this summary suggests, more work is needed before

---

[18]Hermalin and Wallace (1994) and Kaparakis, Miller, and Noulas (1994) found a significant negative relationship; Berger, Hancock, and Humphrey (1993) found a significant positive relationship; and Aly, Grabowski, Pasurka, and Rangan (1990), Berger and Hannan (1996), Cebenoyan, Cooperman, Register, and Hudgins (1993), Mester (1993 and 1996), and Pi and Timme (1993) found an insignificant relationship.

[19]Cebenoyan, Cooperman, Register, and Hudgins (1993), and Hermalin and Wallace (1994) found stock S&Ls more efficient than mutual S&Ls, while Mester (1993) found the reverse, likely because a later sample period was examined.

[20]Mester (1996) found a significant correlation, but Spong, Sullivan, and DeYoung (1995) did not.

[21]Pi and Timme (1993) found a significant negative relationship, Berger and Hannan (1996) found an insignificant negative relationship, and Spong, Sullivan, and DeYoung (1995) found a positive relationship.

[22]See Berger and Hannan (1996).

[23]We do not include financial capital and nonperforming loans in our analysis of correlates described below, since we control for these in the cost and profit models from which our efficiency measures are

a complete picture of financial institution efficiency emerges, and this paper tries to help complete the picture.

5. Empirical Design for Efficiency Estimation

This section outlines and compares the different econometric models used in the estimations below and the assumptions that these models impose on the data. To facilitate exposition and keep the number of comparisons under control, we choose a "preferred" model and measure the effects of deviations from this model one at a time. That is, we choose what we believe to be the best set of variables, best cost and profit function specification, and best frontier efficiency technique within our data and computational constraints and then estimate the effects of making alternative choices one by one, in a controlled experiment. Note that our "preferred" model would not be preferred by most or even necessarily many researchers. There is still substantial disagreement over the best methods of estimation, but they do seem to be converging.

We estimate the efficiency of almost 6,000 U.S. commercial banks that were in continuous existence with complete, accurate data over the six-year period 1990-95.[24] We will discuss here the variables, specification, and estimation method of the preferred model, then briefly mention the alternatives that will be explored in the empirical analysis below.

5.1 Variables Included in the Preferred Specifications of the Cost and Profit Functions. Table 1 gives the definitions of all the variables specified in the cost, standard profit, and alternative profit functions, as well as their sample means and standard deviations for the most recent year of data, 1995. The variable input prices, w, include the interest rates on purchased funds and core deposits as well as the price of labor. Expenditures on these inputs comprise the vast majority of all banking costs. The variable outputs, y, include consumer loans, business loans, and securities, the latter category being measured simply as gross total assets less loans and physical capital, so that all financial assets are considered to be outputs. This specification of financial assets as outputs and financial liabilities and physical factors as inputs is consistent with the

---

derived.

[24]Between 8,378 and 11,077 banks were used in estimating the cost and profit functions each year 1990-95, and the restrictions that all the cost and profit function data be complete for all years left 5,949 banks for which we report efficiencies. These banks had about half of the assets of the U.S. banking industry as of December 1995.

"intermediation" approach or "asset approach" to modeling bank production (Sealey and Lindley, 1977).[25]

We specify risk-weighted off-balance-sheet guarantees, physical capital, and financial equity capital as fixed netputs, z. Off-balance-sheet guarantees are included in the model because they are often effective substitutes for directly issued loans, requiring similar information-gathering costs of origination and ongoing monitoring and control of the counterparties, and presumably similar revenues if these guarantees are competitive substitutes for direct loans. The use of the Basle Accord risk weights implies that these guarantees have approximately the same perceived (according to the Accord) credit risk and, therefore, approximately the same origination, monitoring, and control costs as loans to these same parties. These guarantees are also concentrated in large banks. As a consequence, a scale bias might be present if no account were taken of these items, as larger banks would have disproportionately higher costs relative to their measured outputs. We specify these guarantees as fixed instead of variable primarily because of the difficulty of obtaining accurate output price information for use in specifying the standard profit function.[26]

The treatment of physical capital as a fixed input is relatively standard in efficiency estimation, but specification of equity is not. The reasons for including equity were discussed in section 3.4. As discussed further in section 5.2 below, the specification of equity as fixed helps resolve several estimation problems.

Finally, the environmental variables, v, are limited to the nonperforming loan to total loan ratio either for the bank (NPL) or for the state in which the bank is located (STNPL). In our preferred specification, we use STNPL, since it is almost entirely exogenous and controls for bad luck in the bank's environment. We are not aware of any previous research in which STNPL has been specified. In principle, v could include other measures of the economic conditions faced by the bank, such as the growth or unemployment rate of

[25]In cost function models, deposits are specified as inputs, outputs, or as having both input and output attributes. However, we cannot specify deposits as outputs here, since it would be too difficult to measure an output price for deposits for use in the standard profit function. This is because deposit services are often paid for by paying below-market rates on deposits rather than charging a positive price or fee for services.

[26]A prior study that specified a number of off-balance-sheet activities found that these activities had little effect on cost scale and product mix economies (Jagtiani, Nathan, and Sick, 1995). However, we are unaware of any frontier efficiency studies of either costs or profits that have taken these activities into account.

the state where the bank is located, but these variables are closely related to the state's nonperforming loan record and would make interpretation of the coefficients on NPL and STNPL more difficult. Additionally, we could have included regulatory environmental variables such as state restrictions on branching or on bank holding company expansion. We exclude regulatory information from the efficiency estimates because one of our goals is to test how efficiency is related to these laws by treating them as potential correlates of efficiency.

5.2 Functional Form of the Preferred Specifications. Our preferred model for estimating efficiency specifies the Fourier-flexible functional form, which is a global approximation that includes a standard translog plus Fourier trigonometric terms. For the cost function we specify:

$$
\begin{aligned}
ln\,(C/w_3 z_3) = \alpha &+ \sum_{i=1}^{2} \beta_i\, ln\,(w_i/w_3) + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} \beta_{ij}\, ln\,(w_i/w_3)\, ln\,(w_j/w_3) + \sum_{k=1}^{3} \gamma_k\, ln\,(y_k/z_3) \\
&+ \frac{1}{2}\sum_{k=1}^{3}\sum_{m=1}^{3} \gamma_{km}\, ln\,(y_k/z_3)\, ln\,(y_m/z_3) + \sum_{r=1}^{2} \delta_r\, ln\,(z_r/z_3) + \frac{1}{2}\sum_{r=1}^{2}\sum_{s=1}^{2} \delta_{rs}\, ln\,(z_r/z_3)\, ln\,(z_s/z_3) \\
&+ \sum_{i=1}^{2}\sum_{k=1}^{3} \eta_{ik}\, ln\,(w_i/w_3)\, ln\,(y_k/z_3) + \sum_{i=1}^{2}\sum_{r=1}^{2} \rho_{ir}\, ln\,(w_i/w_3)\, ln\,(z_r/z_3) \\
&+ \sum_{k=1}^{3}\sum_{r=1}^{2} \tau_{kr}\, ln\,(y_k/z_3)\, ln\,(z_r/z_3) + \sum_{n=1}^{7} [\phi_n \cos(x_n) + \omega_n \sin(x_n)] \\
&+ \sum_{n=1}^{7}\sum_{q=n}^{7} [\phi_{nq} \cos(x_n + x_q) + \omega_{nq} \sin(x_n + x_q)] \\
&+ \sum_{n=1}^{7} [\phi_{nnn} \cos(x_n + x_n + x_n) + \omega_{nnn} \sin(x_n + x_n + x_n)] \\
&+ \upsilon_1\, ln\,(STNPL) + \frac{1}{2}\upsilon_{11} [ln\,(STNPL)]^2 + ln\, u_C + ln\, \epsilon_C ,
\end{aligned} \tag{8}
$$

where (y_k/z_3), (z_r/z_3), and the STNPL variables have 1 added for every firm in order to avoid taking the natural log of zero, the x_n terms, n= 1,..., 7 are rescaled values of the *ln* (w_i/w_3), i=l, 2, *ln* (y_k/z_3), k=l, 2, 3, and *ln* (z_r/z_3), r=1, 2, such that each of the x_n is in the interval $[0,2\pi]$, and π refers to the number of radians here (not profits), and the standard symmetry restrictions apply to the translog portion of the function (i.e.,

$\beta_{ij} = \beta_{ji}$, $\gamma_{km} = \gamma_{mk}$, $\delta_{rs} = \delta_{sr}$).[27]

The standard and alternative profit functions use essentially the same specification with a few changes. First, the dependent variable for the profit functions replace *ln* (C/w_3z_3) with *ln* $[(\pi/w_3z_3) + |(\pi/w_3z_3)^{min}| + 1]$, where $|(\pi/w_3z_3)^{min}|$ indicates the absolute value of the minimum value of (π/w_3z_3) over all banks for the same year. Thus, the constant $\theta = |(\pi/w_3z_3)^{min}| + 1$ is added to every firm's dependent variable in the profit function so that the natural log is taken of a positive number, since the minimum profits are typically negative. Thus, for the firm with the lowest value of (π/w_3z_3) for that year, the dependent variable will be $\ln(1) = 0$. For the alternative profit function, this is the only change in specification (other than relabeling the composite error term as *ln* $u_{a\pi}$ + *ln* $\epsilon_{a\pi}$), since the exogenous variables are identical to those for the cost function. For the standard profit function, the terms containing the variable output quantities, *ln* (y_k/z_3), and their trigonometric x_n terms are replaced by the corresponding output prices, *ln* (p_k/w_3), and their x_n trigonometric terms.

The Fourier-flexible form is a global approximation because the cos x_n, sin x_n, cos $2x_n$, sin $2x_n$, etc., terms are mutually orthogonal over the $[0,2\pi]$ interval, so that each additional term can make the approximating function closer to the true path of the data wherever it is most needed.[28] A good fit of the data for the estimated efficient frontier is important in estimating efficiency, because inefficiencies are measured as deviations from this frontier.

As shown in equation (8), all of the cost, profit, input price, and output price terms—including the Fourier terms for prices before transformation—are normalized by the last input price, w_3, in order to impose linear homogeneity on the model. That is, on the efficient frontier, a doubling of all input prices exactly

[27] We cut 10% off each end of the $[0,2\pi]$ interval so that the x_n span $[0.1\times2\pi, 0.9\times2\pi]$ to reduce approximation problems near the endpoints. The formula for x_n is $0.2\pi - \mu\times a + \mu\times$variable, where [a,b] is the range of the variable being transformed, and $\mu \equiv (0.9\times2\pi - 0.1\times2\pi)/(b-a)$. We limit the third-order Fourier terms to include just the interactions of the own terms because of computational limitations in applying the stochastic frontier approach below. The model as shown includes 122 net free parameters after imposing symmetry. We exclude consideration of factor share equations embodying Shephard's Lemma or Hotelling's Lemma restrictions because this would impose the undesirable assumption of no allocative inefficiencies.

[28] The orthogonality is perfect only if the data are evenly distributed over the $[0,2\pi]$ interval, but in practice the Fourier terms have improved the fit of the data in every application of which we are aware.

doubles costs, and a doubling of all input and output prices doubles standard profits.[29] This normalization is the only way to impose homogeneity on the Fourier-flexible specification, since unlike the translog terms, the Fourier terms are not multiplicative.

We specify all of the cost, profit, variable output quantities, and other fixed netput quantities as ratios to the fixed equity capital input, z_3, to control for heteroskedasticity, to help control for scale biases in estimation, and to give the models more economic interpretation. Since the costs and profits of the largest firms are many times larger than those of the smallest firms, large firms undoubtedly would have random errors with much larger variances in the absence of the normalization. In contrast, firms of different sizes have ratios of costs or profits to equity that typically vary only by a few-fold. This is particularly important because the inefficiency terms $ln\ \hat{u}_C$, $ln\ \hat{u}_\pi$, and $ln\ \hat{u}_{a\pi}$ are derived from the composite residuals, which might make the variance of the efficiencies dependent on bank size in the absence of normalization. Similarly, the normalization of the variable output and fixed netput quantities keep these variables from being very skewed for the large banks, so that the dependent and independent variables are roughly of the same order of magnitude.

Normalization by equity also reduces the scale bias discussed in section 2.3 that is likely to be present, particularly in the standard profit function. Large banks will tend to have higher profits for a given set of prices, primarily because they were able to gain size over a period of decades, a feat that small banks cannot achieve in the short run. However, the profits per dollar of equity and assets per dollar of equity of large banks are well within the achievable range for small banks. Moreover, even in the short run, equity is often the variable that limits bank size. Regulators and market participants generally tie the allowable size of the bank, especially its loan portfolio, to its quantity of equity capital available to absorb loan losses. Normalization by equity makes the dependent variable reasonably equally achievable for all banks.

Normalization by equity also has a particular economic meaning. The dependent variable in the profit functions is essentially the return on equity, or ROE, achieved by the bank (normalized by prices and with

---

[29] The homogeneity restriction does not have to be imposed on the alternative profit function, but it is imposed to keep the functional forms equivalent.

a constant added), or a measure of how well the bank is using its scarce financial capital.[30] This measure may be closer to the goal of the bank than maximizing the level of profits, particularly in banking, which is one of the most highly financially levered industries. Shareholders are interested in their rate of return on equity, which is approximated by ROE, and most debtholders do not put much pressure on banks to earn profits because their returns are guaranteed by deposit insurance.

5.3 Preferred Frontier Efficiency Estimation Technique. Our preferred method of estimating efficiency is the distribution-free approach, which disentangles the inefficiency term, $ln\ u$, from the random error term, $ln\ \epsilon$, in equations (2), (4), and (6) by assuming that inefficiencies are relatively stable over time and random errors tend to average out over time.

We briefly sketch the procedure as it is applied here. The cost, standard profit, and alternative profit equations are estimated separately for each year, 1990-95, allowing the coefficients to vary to reflect changes in technology, regulation, and market environment. The average residual for each bank b is formed, which is an estimate of $ln\ u_C^b$, $ln\ u_\pi^b$, or $ln\ u_{a\pi}^b$, depending on the equation. Despite the assumption that random error averages out to zero over time, we realize that the extreme values of these inefficiency estimates may reflect substantial random components. Thus, we use truncation to reassign less extreme values to banks with the most extreme values in each of ten bank size categories. We assign to each bank in the top and bottom 5% of the distribution of the average residuals in a size category the value for the bank that is just at 5th or 95th percentile, respectively. (Other degrees of truncation are also tried, as discussed below.) Truncation is performed within size class deciles (by gross total assets) to reduce the effects of persistently good or bad luck for these banks relative to firms of their size (DeYoung and Nolle, 1996). The resulting estimates of the inefficiency terms, $ln\ \hat{u}_C^b$, $ln\ \hat{u}_\pi^b$, and $ln\ \hat{u}_{a\pi}^b$, along with their minimum or maximum values $ln\ \hat{u}_C^{min}$, $ln\ \hat{u}_\pi^{max}$, and $ln\ \hat{u}_{a\pi}^{max}$, are then substituted into the formulas (5), (6), and (7) above, and the numerators and

[30]Unfortunately, our accounting measure of equity does not perfectly correspond to the market value of the bank, but market values are unavailable for most banks.

denominators are summed over the six years to estimate the efficiency ratios.[31]

The distribution-free method gives a single set of cost, standard profit, and alternative profit efficiency measures for each bank over the entire six-year period, 1990-95. Since it is likely that relative efficiencies among the different banks shift somewhat over time because of changes in management, technical change, regulatory reform, the interest rate cycle, and other influences, this method describes the average deviation of each firm from the best-average-practice frontier. That is, our core efficiency estimates how well a bank tends to do relative to its competitors over a range of conditions over time, rather than a firm's relative efficiency at any one point in time. Besides the fact that this method uses less arbitrary assumptions to disentangle inefficiencies from random error, we believe that by averaging over a number of conditions, this method gives a better indication of a bank's longer-term performance and how it is likely to perform in the future than any method that relies on a bank's performance under a single set of circumstances.

5.4 Deviations from our Preferred Efficiency Measurement Methods. We measure the effects of several deviations from our preferred methods for measuring efficiency to determine the effects of some of the assumptions commonly employed in the efficiency literature. By changing just one assumption at a time, but leaving the data set and all other assumptions unchanged, we aim to isolate the individual effects. The deviations we try are: (1) specifying the translog functional form in place of the preferred Fourier-flexible specification, (2) trying several different specifications of the nonperforming loan ratios (NPL and STNPL), (3) removing equity capital from the model, and (4) using the stochastic frontier approach in place of the distribution-free approach.

6. The Empirical Results Pertaining to the Efficiency Concepts and Measurement Methods

Table 2 shows the means and standard deviations of the efficiencies estimated in the preferred model and in each variation, along with the rank-order correlations of the efficiencies from each variation with those

---

[31]Because the costs and profits in the dependent variables are expressed in terms of ratios to w_3z_3, the $\exp[\hat{f}(\cdot)]$ terms in the efficiency ratios are replaced by $w_3z_3\times\exp[\hat{f}(\cdot)]$, where $\hat{f}(\cdot)$ is the predicted part of the cost or profit function. In order to offset the nonlinearities introduced by exponentiating and including the θ terms, all the predicted costs and profits are multiplicatively adjusted so that the average predicted cost or profit for each year equals the average actual cost or profit for the same year.

from the preferred model. The means and standard deviations are weighted by the denominators of the efficiency ratios (estimated cost or potential profits) to represent the proportion of the entire sample's resources that are used efficiently or potential profits that are earned.[32]

6.1 Efficiency Estimates from the Preferred Model. The mean cost efficiency from the preferred model of 0.868 suggests that about 13.2% of costs are wasted on average relative to a best-practice firm. The 0.868 figure is within the range found in the literature, but is slightly higher than the most typical finding of about 80% cost efficiency. The slightly higher figure might be explained by the fact that we are examining data from the first six years of the 1990s, rather than the 1980s, the period of study for most earlier work.[33]

The mean efficiencies for the standard and alternative profit functions are similar to each other, both showing that about half of the potential profits that could be earned by a best-practice firm are lost to inefficiency. These figures are also well within the observed range from the few other profit efficiency studies. The standard deviations of the profit efficiencies are about 20 percentage points, suggesting that these efficiencies are quite dispersed, with many firms earning considerably more or less than the average figure. By contrast, the cost efficiencies are more tightly distributed with a standard deviation of 6.2 percentage points.

We also note that the alternative profit function does not fit the data nearly as well as the standard profit function. The average adjusted R^2 (not shown) of the cost, standard profit, and alternative profit functions across the six years were 0.931, 0.607, and 0.329, respectively. Apparently, the measured prices of the loan and security outputs are more closely related to the profit dependent variable than are the quantities of these outputs. While a full investigation is beyond the scope of this paper, it seems likely that at least part

---

[32] The cost, standard profit, and alternative profit function coefficient estimates are available upon request from the authors.

[33] The reported efficiency estimates are based on 5% truncation of the average residuals, but other degrees of truncation were also tried. The average measured cost efficiencies at the 0%, 1%, 5%, and 10% truncation levels were 0.689, 0.784, 0.868, and 0.901, respectively. Thus, measured efficiency increases considerably when the degree of truncation rises up to 5%, but the increase tapers off after this point. This suggests that 5% truncation removes most of the random error not already eliminated by averaging over time. Moreover, further truncation would not change the efficiency estimates by any economically meaningful amount. Similar results obtain for the measured profit efficiencies.

of the explanation may lie in differing degrees of service quality or market power. That is, if some banks are providing service qualities that are more in demand and, therefore, are able to charge higher prices, or if some banks are able to exercise market power to raise profits substantially through higher loan prices, this would yield higher explanatory power for the standard profit function.

It is also noteworthy that the alternative profit efficiency ratios are lower on average than the standard profit efficiency ratios. This finding again could be explained by service quality or market power considerations. If, on average, banks are making poor service quality choices relative to the best-practice banks, and these choices are reflected in lower output prices and revenues, the alternative profit efficiency measures would correctly capture this source of inefficiency. Standard profit efficiency is less able to capture the effects of service quality because it takes output prices as given (although they are imperfectly measured). Similarly, if market power in setting output prices tends to explain profitability, the average bank may be further from the alternative profit frontier than from the standard profit frontier, because the alternative frontier does not control for output prices whereas the standard frontier does. Some evidence in favor of these explanations is that the alternative profit efficiency ratio is much more highly correlated with output prices than the standard profit efficiency ratio.[34] Thus, the firms measured as alternative profit inefficient receive relatively low prices for their outputs, perhaps reflecting either low service quality or lack of market power.

6.2 Efficiency Estimates from Variations in Measurement Technique. In the first variation on our preferred model, the translog functional form is substituted for the preferred Fourier-flexible specification by restricting the coefficients of all the trigonometric terms (the x's in equation (8) above) to be zero.[35] The results here suggest only a small difference in average efficiencies and very little difference in efficiency dispersion or rank from using the more restricted specification. The average efficiencies are lower by about

[34] Specifically, the correlations of Alt π EFF with p_1, p_2, and p_3 are 0.288, 0.528, and 0.593, respectively, whereas the corresponding correlations for Std π EFF are 0.198, 0.325, and 0.260, respectively. All are statistically significant at the 1% level.

[35] If one of the specifications fits the data better than the other, it does not necessarily imply that measured efficiency will either increase or decrease. See Berger and DeYoung (1996) for more discussion. They found that the Fourier-flexible specification fit the data better and registered higher measured efficiency, but the efficiencies from both specifications were highly correlated.

1% of costs or potential profits in each case, with about the same degree of dispersion, and rank order correlations of 0.979 or higher with the preferred specification.

Formal statistical tests indicate that the coefficients on the Fourier terms are jointly significant at the 1% level in all 18 cases—the cost, standard profit, and alternative profit functions in all six years. However, the average improvement in goodness of fit or adjusted R^2 is relatively small. Thus, while the null hypothesis of the translog form is rejected from a statistical viewpoint and the Fourier-flexible efficiency estimates are likely more accurate, the improvement in fit is not significant from an economic viewpoint. Both functional forms yield essentially the same average level and dispersion of measured efficiency, and both rank the individual banks in almost the same order.

The next three variations in the table contain alternative specifications of nonperforming loans. In contrast to our preferred model, which specifies the state's nonperforming loan ratio in first- and second-order logged terms [ln STNPL and ½ $(ln \text{ STNPL})^2$], the next three specifications: (1) replace these terms with the bank's own ln NPL and ½ $(ln \text{ NPL})^2$; (2) include all of the STNPL and NPL terms; and (3) include none of the STNPL or NPL terms. The efficiency estimates are strikingly similar across the four specifications (including the preferred specification). The average efficiencies are all within 1 percentage point of each other, the measured dispersion is virtually identical, and rank-order correlations are all over 99%. Apparently, given the rest of our specification, the treatment of nonperforming loans is not materially important to the efficiency estimates.

Nonetheless, the coefficients of the STNPL and NPL variables in the cost and profit functions (not shown) did yield some insights as to which of the main hypotheses about the effects of nonperforming loans—"bad luck," "bad management," or "skimping"—is most consistent with the data. For each of the cost and profit equations, we formed the derivatives of the dependent variable with respect to STNPL and/or NPL evaluated at the mean values of the data (not shown). The data primarily supported the "bad management" hypothesis—i.e., that firms that are inefficient at managing their operations are also poor at managing their loan portfolios. In almost every case, the derivative with respect to NPL was unfavorable (positive for costs, negative for profits) and statistically significant at the 1% level. That is, firms with loan performance problems also tended to have high costs and low profits, consistent with the “bad management” hypothesis. This occurred whether or not STNPL was specified in the same equation, which should remove much of any “bad luck” effect. The derivative with respect to STNPL was often statistically significant in the predicted direction at the 1% or 5% level, but was not as consistent as the results for NPL. Thus, the “bad luck” hypothesis—under which exogenous conditions cause loan performance problems that raise costs—received more limited support than the “bad management” hypothesis. The “skimping” hypothesis—under which nonperforming loans may be associated with **low** costs from choosing to put less effort into loan monitoring and control—was generally not supported by the data or its consequences were overwhelmed by “bad management” effects. These results support our choice of specifying STNPL in the cost and profit functions and excluding NPL. STNPL appears to be a useful control variable for the bank’s economic environment and is almost completely exogenous, whereas the inclusion of NPL likely adds an endogenous efficiency factor that should not be controlled for when estimating efficiency.

In the next variation, the equity capital fixed input z_3 is eliminated from the cost and profit functions. There is little effect on the average level or dispersion of cost efficiency, although the firms are ranked slightly differently, with a rank order correlation of 0.834 with the preferred model. More important, the average profit efficiencies fell from means of about 50% of potential profits earned on average to about 10%, with a much higher standard deviation and much lower rank-order correlation with the estimates from the preferred specification. This is not unexpected. As discussed above, the specification of equity as a fixed input in the preferred model reduces the scale bias that may be created by the fact that the equity capital of small banks cannot be expanded to match that of large banks and allow them to expand their asset portfolios greatly except after a period of decades. The dependent variable in this variation depends on the level of profits, which can be much higher for large banks, as opposed to our preferred specification, where the dependent variable is a function of the rate of return on equity, which is more comparable across size classes.

A breakdown of the measured profit efficiencies by size class when equity is removed (not shown) is consistent with these arguments. Banks with gross total assets below $50 million had mean measured standard and alternative profit efficiencies of –0.068 and 0.020, respectively, whereas the corresponding efficiencies for banks with over $10 billion were 0.768 and 0.783. Clearly, the removal of the equity control variable rewards large banks that have high levels of profits by virtue of their equity positions that have been built up over time, but these firms generally do not have particularly high rates of return on their equity. The evidence from this variation strongly supports our specification of equity capital as fixed in the preferred model.

The final variation shown in Table 2 uses the stochastic frontier approach where the inefficiencies are assumed to be half-normally distributed. The data, however, do not appear to fit that distribution very well. As described in the table, the skew of the data was not consistent with the half-normal assumptions in a number of cases. As a result, we do not have any standard profit efficiency estimates, and the cost and alternative efficiencies must be based on partial samples.

Despite these difficulties, the efficiencies estimated by the stochastic frontier approach are reasonably consistent with those of the preferred distribution-free approach. The average cost efficiency is somewhat higher with less dispersion, and the average alternative profit efficiency is somewhat lower with more dispersion, but in both cases, the rank-order correlations with the distribution-free method are over 90%.

Overall, the results of Table 2 suggest that variations in methodology usually do not affect measured efficiency substantially, except that profit efficiencies may be significantly scale biased when equity capital is excluded. For the most part, the findings support the choices made in our preferred approach and give us confidence to proceed from this point forward only with the estimates derived from the preferred model.

6.3 Cost Scale Economies in the Preferred Model. Table 3 shows cost ray scale efficiency by size class and breaks out the X-efficiency by size class as well. Scale efficiency is defined as the ratio of predicted minimum average costs to actual average costs, both adjusted to be on the X-efficient frontier, (i.e., setting $ln\ u_C$ to the minimum in the sample and $ln\ \epsilon_C$ to zero). Total cost efficiency is the product of the scale and

X-efficiency ratios.[36] The variable t^* is the ratio of efficient scale to actual scale.

The basic result shown in Table 3 is that in every size class, the typical bank shows unexploited ray scale economies—i.e., that the bank's product mix could be produced at lower average cost by increasing the scale of output. The mean scale efficiencies are around 80%, suggesting that approximately equal amounts of resources are lost because of scale and X-inefficiencies. In every size class more than 90% of firms are operating below efficient scale, and the mean t* is between *2* and 3 for each size class, suggesting that the typical bank would have to be 2 to 3 times larger in order to maximize cost scale efficiency for its product mix and input prices.[37]

To ensure that our scale economies estimates reflect the shape of the frontier and are not simply the consequence of correlation between the X-inefficiencies and scale, we reestimated the Fourier-flexible cost model using only relatively efficient banks. We divided the banks into asset size deciles, chose the top 25% of the distribution in terms of efficiency scores in each size class, reestimated the cost frontier model using these banks, and then recomputed the scale measures. We got very similar results to those shown in Table 3.[38]

These findings differ from most of those found using 1980s data, in which large banks were typically found to be operating at constant returns to scale or with slight cost diseconomies of scale. In almost all cases, cost scale economies were exhausted well below $10 billion in assets .[39] The difference could have occurred because of some of the differences in specification between our cost function and those typically employed in the literature. One candidate is our use of the Fourier-flexible function form in place of the more common

[36]The standard profit efficiencies are already inclusive of scale efficiency. We do not compute scale efficiencies for the alternative profit function because they would include economies on the consumer side, and would not be comparable (see Berger, Humphrey, and Pulley, 1996).

[37]As shown in Table 3, unexploited scale economies are actually somewhat larger for banks in the larger size classes. This would be consistent with the hypothesis that larger banks choose product mixes that are more conducive to large scale.

[38]The weighted average scale efficiency across all banks in this estimation was 0.815, the weighted average t^* was 1.854 (significantly greater than 1), and a weighted average of 93.5% of bank costs were incurred by banks that were operating below efficient scale.

[39]Exceptions include Hunter and Timme (1986), Shaffer and David (1991), Shaffer (1994), Hughes, Lang, Mester, and Moon (1995, 1996a), and Hughes and Mester (1996).

translog specification. The translog forces a symmetric U-shape in logs on the ray average cost frontier, which could force measured scale diseconomies on the large banks as the imposed reflection of the scale economies found for small banks. However, this does not appear to be the case here. We reestimated the scale economies using the translog (not shown) and found that, if anything, the measured scale economies for the larger banks were even greater. Another candidate is our use of equity capital. Failure to control for equity could give a bias toward finding cost scale diseconomies, because large banks tend to have lower equity ratios and pay interest on higher portions of their funds. Reestimation without specifying equity capital (not shown) did reduce our measured scale economies for large banks, consistent with the expected bias, but it did not eliminate them. The last column of Table 3 reports the results of another robustness check of these scale economy findings. We examined the raw data without imposing a specification. The ratio C/GTA, average cost per dollar of assets, falls consistently when moving into larger size classes. This simple measure is, if anything, biased against the larger banks, which typically have more off-balance-sheet guarantees and more loans per dollar of assets, which should raise average costs. The finding of declining cost per dollar of assets by size class strongly supports our scale economy findings. Moreover, an examination of the ratio C/GTA for banks over the 1980s reveals mild scale economies for asset levels below $1 billion and diseconomies for larger banks, which is consistent with the prior scale economies literature. This suggests that the 1990s are indeed different, and that our methodology has not created the result.

An important caveat to the scale efficiency findings is that they may not hold for the very largest banks. We group the data from all banks with GTA greater than $10 billion into a single size class in Table 3 because there are too few very large banks in the U.S. to form credible size subclasses of this largest size class, and our data exclusion rules exacerbated the problem by dropping several of these banks. To try to ameliorate this problem, we recalculated the C/GTA ratio in the last column of Table 3 including all U.S. banks (regardless of our data exclusion rules), and segmented the largest size class into $10 billion -$25 billion, $25 billion -$50 billion, and above $50 billion ranges. We find average costs to be decreasing in all size classes up through $25 billion, with an increase in average costs thereafter. Thus, we still find relatively robust evidence of scale economies well beyond the region usually found in studies using the 1980s data. Serious estimates of scale economies for U.S. banks over $25 billion will likely have to wait for the consolidation of the industry to create enough of these large banks to yield reasonable estimates.

Our scale economy results suggest that some conditions changed between the 1980s and 1990s that substantially raised the cost-efficient scale of U.S. banks. While a complete investigation is beyond the scope of this paper, three explanations seem plausible. First, open-market interest rates have been relatively low recently—the one-year U.S. Treasury rate averaged 9.74% in the 1980s and 5.39% over our sample period of 1990-1995. It is quite likely that these low rates reduced interest rate expenses (which account for most of costs) proportionally more for large banks than small banks, because a greater proportion of large banks' liabilities tend to be market-sensitive. Large banks often rely on wholesale purchased funds that pay market rates, whereas small banks typically rely more on core deposits with rates that do not vary one-for-one with open-market rates. Under this explanation, the scale economies of the early 1990s may be a temporary phenomenon that will disappear if and when market rates rise substantially. To partially check this explanation, we reevaluated each bank's scale economies as if interest rates were at their levels in the 1980s.[40] We still find scale economies (although they are slightly lower), suggesting that a rise in interest rates back to the higher levels of the 1980s would not eliminate the scale economies. We acknowledge, however, that our interest-rate experiment is subject to the Lucas critique. Namely, the cost function was estimated for the low interest rate environment of the 1990s. Because rates are low, banks might be paying less attention to optimizing with respect to interest rates. As interest rates rise, the parameters of the cost function could change in a way that affects scale economies significantly.

A second possible explanation for our scale economies result is that recent regulatory changes may have tended to favor large banks relative to small banks. In particular, the elimination of geographic

---

[40] In particular, for each bank in the sample that had at least nine years of data for the 1980s, we calculated the average interest rate the bank paid on purchased funds and core deposits in the 1980s and the average interest rates it paid for these funding sources in the 1990s, and found the differences in average rates paid. We then added each bank's difference to its core deposit rate in each year in the 1990s and recalculated the bank's scale economies at this new evaluation point. On average, this added 2.55% to the purchased funds rate and 2.50% to the core deposit rate.

restrictions on bank branching and holding company expansion during the 1980s and into the early 1990s may have removed some scale diseconomies and made it less costly to become large. For example, in the extreme case of unit banking, there are very severe diseconomies to becoming large without being able to have any branch offices to collect deposits, and such diseconomies would be removed by allowing statewide branching.[41] Similarly, the removal of interest rate ceilings on core deposits during the 1980s likely raised costs more for small banks, which rely more on core deposits for their funding.

Finally, improvements in technology and applied finance may have reduced costs more for large banks than for small banks. Improvements in information processing and credit scoring may have reduced the costs of extending small business loans and credit card loans more for larger banks. Similarly, improved automation may have allowed large banks to expand faster and at lower cost by setting up ATM machines in place of adding more expensive brick-and-mortar branch offices. Large banks may have also been better positioned to take advantage of the new tools of financial engineering, such as derivative contracts and other off-balance-sheet activities.

6.4 Comparison of Efficiency Across Concepts. Table 4 shows the rank-order correlations among the different X-efficiency measures and some other commonly used financial ratios that may be considered raw-data measures of efficiency. Standard and alternative profit efficiency are highly positively and statistically significantly correlated with each other ($\rho = 0.794$), as expected. Perhaps surprisingly, however, measured cost efficiency is essentially uncorrelated with standard profit efficiency and it is negatively correlated with alternative profit efficiency.

One possible explanation is that cost and revenue inefficiencies may be negatively related, so that firms with low cost efficiency tend to have high revenue efficiency that offsets it. This could occur because of competitive pressures if, for example, firms with highly valued product mixes or high revenue efficiency feel less market discipline to control their costs.

[41] Jayaratne and Strahan (1996) found that bank performance measured several different ways improved after within-state branching restrictions were removed, although they did not separate the improvement into scale and X-efficiency effects.

An alternative explanation is that much of what are measured as cost inefficiencies are actually unmeasured differences in product quality that required additional costs to create. As discussed above, the alternative profit function and to a lesser extent the standard profit function tend to control for product quality implicitly by letting revenues received for higher quality offset the extra costs of creating the quality. We will explore this possibility more in the next table, where we compare cost and profit inefficiencies.

The correlations between the efficiencies and each of the raw-data measures follow the expected pattern—efficiency by any definition is negatively and significantly correlated with the standard average cost ratio C/GTA and positively and significantly correlated with the standard profitability ratios ROA and ROE. These findings suggest that our efficiency measures are robust and are not simply the consequences of our specifications or methods.

As measured, the cost and profit efficiency ratios are not directly comparable because they are reported in terms of different denominators (predicted actual costs versus potential profits). In Table 5, we report the dollar values of the cost, standard profit, and alternative profit inefficiencies divided by the same denominators—potential profits, gross total assets, and equity. The meanings of these ratios are the proportions of potential profits lost, the loss of ROA, and the loss of ROE, respectively, because of inefficiency. These ratios may provide evidence on the extent to which the measured cost inefficiencies incorporate unmeasured differences in product quality. If markets are competitive so that differences in product quality are rewarded with higher revenues that cover the costs, the alternative profit inefficiency essentially just improves on cost efficiency by offsetting the extra costs of producing higher quality with higher revenues. In this event, the alternative profit inefficiency ratios measured here would be expected to be smaller than the cost inefficiency ratios. If instead, the effects of market power in pricing bank outputs are more important, alternative profit inefficiency may be larger than both cost inefficiency and standard profit inefficiency, as firms with the most market power are measured as much more alternative-profit efficient than the average bank.

The empirical results suggest that this market power paradigm appears to dominate any effects of unmeasured differences in product quality on measured efficiency. Alternative profit inefficiency is larger than are both cost and standard profit inefficiencies. This does not suggest that unmeasured differences in product quality are unimportant, just perhaps less important than market power considerations in determining bank profits. These results also suggest that the standard assumption of perfect competition in setting output prices maintained for measuring standard profit efficiency may be violated by the data.

7. Empirical Investigation of the Potential Correlates of Efficiency

The last part of our analysis relates our efficiency estimates to various aspects of the banks, their markets, and their regulation that are potential correlates of efficiency, i.e., factors that are at least partially exogenous and may explain some of the efficiency differences that remain after controlling for efficiency concept and measurement method. Here, we use the three distribution-free X-efficiency measures estimated for the 1990-95 period and the average values of the bank, market, and regulatory characteristics over 1990-95. As we did when estimating efficiency, we also investigate a few alternative specifications of the potential correlates to check robustness.

The characteristics we investigate are given in Table 6, where we show their definitions, means and standard deviations over 1990-95. These variables fall into six broad categories: **bank size, organizational form and corporate governance, other bank characteristics, market characteristics, state geographic restrictions on competition,** and **primary federal regulator.**

We performed both multiple regressions and single variable regressions. Including an endogenous variable in a multiple regression can bias the coefficients even on the exogenous variables, and perhaps all of our variables are partly endogenous and partly exogenous. Thus, in addition to the multiple regressions, we also ran regressions that each included a constant term and a single explanatory variable. These single-variable regression coefficients are proportional to correlation coefficients. A disadvantage of these single variable regressions is that any significant correlation found might be spurious, with both efficiency and the included variable being significantly related to a third, omitted factor. Because the multiple and single regression analyses each has advantages and disadvantages, we will be conservative and tend to draw conclusions only when the coefficients in both are statistically significant and of the same sign. These results are shown in Table 7.

7.1 Bank Size. The bank size variables (SMLBANK, MEDBANK, LARBANK, HUGBANK) are measured with dummy variables to allow for nonmonotonicity and nonlinearities in the relationship between bank size and efficiency. We specified our multiple regression equation without a constant term and included all four of these size variables.[42] As shown in Table 7, the cost efficiency estimates do not vary much across size classes—holding all else equal, the cost efficiency is about 2.5% higher at the largest banks (with assets over $10 billion) than the smallest banks (with assets under $100 million). But in terms of profit efficiency (both standard and alternative), small banks show the greatest level of efficiency. This result suggests that our profit efficiency measures display very little of the potential scale biases favoring larger firms discussed above (as long as equity capital is specified). The cost and profit efficiency results together seem to imply that as banks grow larger, they are equally able to control costs, but it becomes harder to efficiently create revenues. This is consistent with conventional wisdom and the historical fact that small banks typically have higher profitability ratios. It also helps explain the lack of a positive correlation between cost efficiency and profit efficiency discussed above.

7.2 Organizational Form and Corporate Governance. The banking industry is consolidating at a rapid pace, so it is important to determine the efficiency effects of bank mergers and acquisitions. Our results indicate that, all else equal, banks that have survived at least one merger over our sample period (MERGED) have higher standard and alternative profit efficiency than other banks, which is consistent with prior findings discussed above. This result is not confirmed, however, in the single variable regressions, where alternative profit efficiency is significantly lower for banks that have been involved in mergers. This is perhaps because being involved in a merger is correlated with bank size, and size is negatively related to alternative profit efficiency. On the other hand, being acquired by another holding company (ACQUIRED) does not appear to be associated with profit efficiency, but is associated with higher cost efficiency. All in

---

[42] We excluded one of the state branching dummy variables (STATEB) and one of the primary federal regulator dummy variables (OCC) to avoid perfect collinearity in the regressions.

all, these results are fairly mixed.

Banks with complicated organizational forms or internal management structures could be less efficient, but a holding company structure might also impose some discipline on banks, so we explored the relationships between efficiency and whether the bank is in a bank holding company (INBHC), whether the holding company is multilayered (MUL_LAY), and whether the top-tier holding company is located out of state (OUTST), which could make control more difficult. Banks in holding companies tend to have higher levels of profit efficiency (both standard and alternative) than independent banks, and their cost efficiency is significantly greater as well, consistent with some previous cost efficiency studies. If the holding company has multiple layers, this means even higher levels of profit and cost efficiency. Thus, the more complex structure of multilayered holding companies does not appear to be harming bank efficiency. Having the highest holding company owner from out of state also is associated with higher, not lower, cost efficiency. A potential explanation for these results may be a form of the efficient structure hypothesis (Demsetz, 1973)—more efficient banking organizations may tend to acquire other banks, and the multilayer, multistate holding company is the vehicle that allows them to do it.

We also included a variable indicating whether the bank's highest holder is registered with the SEC for public trading (PUB_TRADED). To the extent that outside shareholders can exert control over bank management, we might expect publicly traded banks to be more efficient, all else equal, and this is indeed what our results indicate. Publicly traded firms tend to have both higher cost and standard profit efficiencies.

For 126 of the banks we have information from 1987-88 on the proportion of stock owned by insiders, i e., board members and their relatives (INSIDE) and the proportion of stock owned by outsiders who had more than 5% of the outstanding shares (OUTSIDE).[43] So in an alternative specification with the more limited number of observations (not shown in Table 7), we also included the first- and second-order terms of these two variables [INSIDE, ½ $INSIDE^2$, OUTSIDE, and ½ $OUTSIDE^2$]. We included the second-order terms because Gorton and Rosen (1995) predict a U-shaped relationship between insiders' stock holdings and

[43]We thank Gary Gorton and Rich Rosen for providing these data.

efficiency. At low levels of inside ownership, a negative relationship would be consistent with managers who have greater control pursuing their own interests, which may involve inefficiencies. However, at higher stakes, an increase in insider ownership may serve to align management's objectives with those of owners, yielding greater efficiency. We include the OUTSIDE variables, since outside investors can be a controlling influence, and the more stock in the hands of large, outside investors, the more control these investors can be expected to exert and the more efficient the bank might become. However, in none of our regressions are any of these governance variables significantly related to efficiency.

7.3 Other Bank Characteristics. We also explored the effects of other characteristics of the bank. A bank's age (AGE) might be related to efficiency since bank production might involve "learning by doing" (Mester, 1996). While significantly different from zero in our profit efficiency regressions, the coefficient on AGE is very small in all the regressions.

We included several variables to control for the strategic niche of the bank, including proxies for the amount of risk the bank is taking on. Banks with higher loan-to-asset ratios (LOAN/GTA) tend to have higher profit efficiency. This might reflect that banks' loan product is more highly valued than securities, or it could reflect higher market power that exists in loan markets compared to the other product markets in which banks operate. Whether a bank is heavily using derivative contracts, such as swaps, forwards, and futures (SUB_DER), does not appear to be consistently related to its efficiency. This might be because of heterogeneity in the uses of these instruments, which can be used for both hedging and speculative purposes. Reliance on purchased funds (PF/GTA) could also be related to efficiency, since the cost of purchased funds differs from that of core deposits over the business cycle. We find that banks that use more of these funds tend to have lower profit efficiencies than other banks.

As a direct measure of bank risk, we included the standard deviation of return on assets (SDROA). To the extent that we are not adequately controlling for risk taking in our profit models, riskier banks may be more profit efficient if they are trading off between risk and return. Alternatively, banks that are poor at operations might also be poor at risk management, which would imply a positive relationship between profit and/or cost efficiency and risk. The evidence suggests that banks with more variable returns tend to have lower profit efficiencies (in the multiple regressions) and also lower cost efficiencies, consistent with the notion that bad managers are poor at both operations and risk management. A negative relationship with cost efficiency was also found when we replaced SDROA with the standard deviation of the return on book equity (SDROE) as a robustness check.

7.4 Market Characteristics. The next set of variables characterizes the competitive conditions of the markets in which the banks operate. The Herfindahl index (HERF) measures the degree of local deposit market concentration and proxies for the bank's market power. As might be expected, market power is negatively related to cost efficiency but positively related to alternative profit efficiency. Banks in less competitive markets can charge higher prices for their services but might feel less pressure to keep costs down (i.e., enjoy the "quiet life").[44] We included two other variables related to market competition. INMSA indicates whether the bank is located in a metropolitan area, which may be more competitive than a rural area. The results are not consistent across the multiple and single variable regressions, because INMSA is likely correlated with other aspects of the bank, like its size, merger activity, etc. State income growth (STGROW) proxies for the growth of market demand for banking services. Greater demand might allow for less cost efficient production, at least in the short run before new competitors enter, but more profit efficiency, since it means greater opportunity to make profit. This is what our estimates show.[45]

7.5 State Geographic Restrictions on Competition. Geographic restrictions on bank expansion, which differ across states, can also limit the competitive forces banks face. We included variables to control for the degree of branching restrictions (UNITB, LIMITB, STATEB), the degree of in-state holding company expansion permitted (LIMITBHC), whether out-of-state holding company expansion is allowed or not

---

[44]However, this result is not robust to replacing HERF with SHARE, the bank's share of local market deposits, as SHARE is significantly positively related to cost efficiency.

[45]As a robustness check we replaced STGROW with the state unemployment rate (STUNEMP), which is negatively related to market demand. Here the results are much weaker, and the signs on STUNEMP in the profit regressions are positive, indicating a negative relationship between market demand and profit efficiency.

(NOINTST), and the proportion of the banking industry's assets held in any states that are allowed to enter (ACCESS). These variables are meant to control for the degree of competition or contestability of the bank's market. Perhaps the most surprising finding here is that the relationship between branching restrictions and efficiency does not appear to be monotonic in the severity of the restrictions. We find some support that banks in states with limited branching restrictions have higher cost efficiency and lower profit efficiency than banks in either unit banking states or states without branching restrictions; banks in unit banking states appear to be the least cost efficient but most profit efficient. The other variables included to measure geographic restrictions produce no consistently significant results, except that banks in states that prohibit interstate expansions (NOINTST) appear to be less cost efficient than banks in states that permit such expansions. In summary, efficiency does seem to be related to limits on geographic expansion, but the findings are somewhat inconsistent.

7.6 Federal Bank Regulator. Our last group of variables—the identity of the bank's primary federal regulator—helps account for some control for the regulatory regime banks are facing. The regulator, like the market, exerts some control over the bank and thus might be related to bank efficiency. Also, the bank's primary federal regulator (FED, OCC, or FDIC) varies depending on the type of charter the bank has; thus, the variables might reflect differences in banks with different charters. We find only weak relationships between regulator identity and our three efficiency measures, with banks overseen by the Federal Reserve tending to be more cost efficient than banks overseen by the OCC or FDIC.

7.7 Fit of the Correlates Equations. A final observation from Table 7 is that the adjusted R^2s suggest that we have not explained most of the variance in measured efficiency. Our 25 explanatory variables are able to explain about 7% of the variance of measured cost efficiency and about 35% of the variance of the two types of measured profit efficiency. We make no judgment as to whether these figures are high or low, but simply note that most of the variance in measured efficiency after controlling for efficiency concept and measurement methods remains unexplained. It may be due to essentially unmeasurable factors, such as differences in managerial ability, to potential correlates of efficiency that we could have included but failed to do so, or to measurement error in the dependent variable due to the many difficulties in measuring efficiency. Further investigation is beyond the scope of this paper.

8. Conclusion

Despite the very significant research effort that has been mounted over the last few years examining the efficiency of financial institutions, there is as yet little information and no consensus on the sources of the substantial variation in measured efficiency, i.e., these sources remain a “black box.” Here, we focus on getting inside the box by examining a number of sources, holding the data set constant. We examine three types of sources: (1) differences in the efficiency concept used, (2) differences m efficiency measurement methodology within the context of these concepts, and (3) the potential correlates of efficiency that may explain some of the efficiency differences that remain after controlling for efficiency concept and measurement method. We review the literature on the sources of efficiency at commercial banks and provide new evidence using a large data set of almost 6,000 U.S. commercial banks that were in continuous operation over the six-year period 1990-95.

We examine three economic efficiency concepts—cost, standard profit, and alternative profit efficiencies. Each corresponds to how well a firm performs relative to a different economic optimization program, and so each may provide different insights about firm efficiency. Consistent with this expectation, we find that measurement of each of the efficiency concepts does add some independent informational value. In fact, the measures of profit efficiency are not positively correlated with cost efficiency, even though all three efficiency measures are positively related to some raw-data measures of performance. As well, a number of the potential correlates had different relationships with the three different efficiency measures, again suggesting that each is measuring a different type of optimization. These results suggest that future researchers might consider measuring all three concepts to be sure that any conclusions about which firms are most efficient or which potential correlates succeed in “explaining” efficiency are robust with respect to all three economic efficiency concepts.

We explore the effects of a number of different efficiency measurement methods on each of the three efficiency concepts. These methods include the use of different measurement techniques, different functional forms, and various treatments of output quality and financial capital. The results for each of the efficiency concepts are quite robust. We find that the choices made concerning measurement technique, functional form, and other variables usually make very little difference in terms of either average industry efficiency or the rankings of individual firms in our data set. An exception is the treatment of equity capital. Failure to account for the equity position of a bank seems to yield a strong scale bias, making large banks appear to be more efficient than small banks by virtue of the equity they have built up over time.

We also find substantial unexploited cost scale economies for our 1990s data up to bank sizes much larger than typically found in the past. This might have occurred because the decline in interest rates, regulatory changes such as the liberalization of intrastate and interstate banking, and improvements in technology and applied finance since the 1980s may have tended to favor large banks over small banks.

Our analysis of the potential correlates of bank efficiency cover a number of bank, market, and regulatory characteristics using multiple- and single-variable regressions. The results are quite mixed. Some of the potential correlates of efficiency have the predicted sign and statistical significance; others have little independent influence on efficiency; and some have unexpected or mixed signs. Importantly, most of the variance in measured efficiency for each of the efficiency concepts remains unexplained, because of unmeasured factors such as differences in managerial ability, potential correlates that were inadvertently excluded from the analysis, or measurement error in the efficiency dependent variables. We leave to future research the task of better explaining efficiency or determining that not much more can be explained.

We close with a caveat that the empirical results of this study should not be taken too seriously unless confirmed by future research. To our knowledge, this is the first study of the efficiencies of the first six years of the 1990s, the first to compare cost, standard profit, and alternative profit efficiency of banks using a single data set and consistent specifications, the first to evaluate the effects of so many differences in methodology, and the first to use such a comprehensive set of potential correlates of efficiency. It is quite likely that some of these results will differ in future studies of these data that use different sets of assumptions.

References

Akhavein, J.D., A.N. Berger, and D.B. Humphrey, 1997, The effects of megamergers on efficiency and prices: Evidence from a bank profit function, Review of Industrial Organization 12.

Akhavein, J.D., P.A.V.B. Swamy, and S.B. Taubman, 1997, A general method of deriving the efficiencies of banks from a profit function, Journal of Productivity Analysis 8.

Aly, H.Y., R. Grabowski, C. Pasurka, and N. Rangan, 1990, Technical, scale, and allocative efficiencies in U.S. banking: An empirical investigation, Review of Economics and Statistics 72, 211-218.

Bauer, P. W. and D. Hancock, 1993, The efficiency of the Federal Reserve in providing check processing services, Journal of Banking and Finance 17, 287-311.

Berg, S.A., F.R. Førsund, and E.S. Jansen, 1992, Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980-89, Scandinavian Journal of Economics 94, S211-S228.

Berger, A.N., 1993, "Distribution-free" estimates of efficiency in the U.S. banking industry and tests of the standard distributional assumptions, Journal of Productivity Analysis 4, 261-292.

Berger. A.N., D. Cummins, and M. Weiss, 1996, The coexistence of multiple distribution systems for financial services: The case of property-liability insurance, Working paper, Wharton Financial Institutions Center, University of Pennsylvania.

Berger, A.N. and R. DeYoung, 1996, Problem loans and cost efficiency in commercial banks, Working paper, Board of Governors of the Federal Reserve System.

Berger. A.N., D. Hancock, and D.B. Humphrey, 1993, Bank efficiency derived from the profit function, Journal of Banking and Finance 17, 317-347.

Berger, A.N. and T.H. Hannan, 1989, The price-concentration relationship in banking, Review of Economics and Statistics 71, 291-299.

Berger, A.N. and T.H. Hannan, 1996, The efficiency cost of market power in the banking industry: A test of the "quiet life" and related hypotheses, Working paper, Federal Reserve Board.

Berger, A.N. and D.B. Humphrey, 1991, The dominance of inefficiencies over scale and product mix economies in banking, Journal of Monetary Economics 28, 117-148.

Berger, A.N. and D.B. Humphrey, 1992, Measurement and efficiency issues in commercial banking, in: Z. Griliches, ed., Measurement issues in the service sectors (National Bureau of Economic Research, University of Chicago Press, Chicago) 245-279.

Berger, A.N. and D.B. Humphrey, 1997, Efficiency of financial institutions: International survey and directions for future research, European Journal of Operational Research.

Berger, A.N., D.B. Humphrey, and L.B. Pulley, 1996, Do consumers pay for one-stop banking? Evidence from an alternative revenue function, Journal of Banking and Finance 20, 1601-21.

Berger, A.N., J.H. Leusner, and J. Mingo, 1996, The efficiency of bank branches, Working paper, Board of Governors of the Federal Reserve System.

Cebenoyan, A.S., E.S. Cooperman, C.A. Register, and S.C. Hudgins, 1993, The relative efficiency of stock vs. mutual S&Ls: A stochastic frontier approach, Journal of Financial Services Research 7, 151-170.

Clark, J., 1996, Economic cost, scale efficiency and competitive viability in banking, Journal of Money, Credit, and Banking, 28 (August): 342-64.

Demsetz, H., 1973, Industry structure, market rivalry, and public policy, Journal of Law and Economics, 16: 1-9.

DeYoung, R., 1997, A diagnostic test for the distribution-free efficiency estimator: An example using U.S. commercial bank data, European Journal of Operational Research.

DeYoung, R. and D. Nolle, 1996, Foreign-owned banks in the U. S.: Earning market share or buying it? Journal of Money, Credit, and Banking 28.

English, M. and K. Hayes, 1991, A simple test of market power, Working paper, Southern Methodist University.

Fried, H.O., C.A.K. Lovell, and P.V. Eeckaut, 1993, Evaluating the performance of U.S. credit unions, Journal of Banking and Finance 17, 251-265.

Gardner, L.A. and M.F. Grace, 1993, X-efficiency in the U.S. life insurance industry, Journal of Banking and

Finance 17, 497-510.

Gorton, G. and R. Rosen, 1995, Corporate control, portfolio choice and the decline of banking, Journal of Finance 50, 1377-1420.

Greene, W.H., 1990, A gamma-distributed stochastic frontier model, Journal of Econometrics 46, 141-163.

Hancock, D., 1985, The financial firm: Production with monetary and nonmonetary goods, Journal of Political Economy 93, 859-880.

Hancock, D., 1986, A model of the financial firm with imperfect asset and deposit elasticities, Journal of Banking and Finance 10, 37-54.

Hannan, T.H., 1991, Bank commercial loan markets and the role of market structure: Evidence from the survey of commercial lending, Journal of Banking and Finance 15, 133-149.

Hannan, T.H. and N. Liang, 1993, Inferring market power from time-series data: The case of the banking firm, International Journal of Industrial Organization 11, 205-218.

Hermalin, B.E. and N.E. Wallace, 1994, The determinants of efficiency and solvency in savings and loans, Rand Journal of Economics 25, 361-381.

Hughes, J.P., W. Lang, L.J. Mester and C.-G. Moon, 1995, Recovering technologies that account for generalized managerial preferences: An application to non-risk-neutral banks, Working paper no. 95-8/R, Federal Reserve Bank of Philadelphia.

Hughes, J.P., W. Lang, L.J. Mester and C.-G. Moon, 1996a, Efficient banking under interstate branching, Journal of Money, Credit, and Banking 28.

Hughes, J.P., W. Lang, L.J. Mester and C.-G. Moon, 1996b, Safety in numbers? Geographic diversification and bank insolvency risk, Working paper 96-14, Federal Reserve Bank of Philadelphia.

Hughes, J. and L.J. Mester, 1993, A quality and risk-adjusted cost function for banks: Evidence on the "too-big-to-fail" doctrine, Journal of Productivity Analysis 4, 293-315.

Hughes, J. and L.J. Mester, 1996, Bank capitalization and cost: Evidence of scale economies in risk management and signaling, Working paper 96-2, Federal Reserve Bank of Philadelphia.

Hughes, J. and C.-G. Moon, 1995, Measuring bank efficiency when managers trade return for reduced risk, Working paper, Department of Economics, Rutgers University.

Humphrey, D.B. and L.B. Pulley, 1997, Banks' responses to deregulation: Profits, technology, and efficiency, Journal of Money, Credit, and Banking 29.

Hunter, W.C. and S.G. Timme, 1986, Technical change, organizational form, and the structure of bank production, Journal of Money, Credit, and Banking 18, 152-166.

Jagtiani, J., A. Nathan, and G. Sick, 1995, Scale economies and cost complementarities in commercial banks: On- and off-balance sheet activities, Journal of Banking and Finance 19, 1175-1189.

Jayaratne, J. and P.E. Strahan, 1996, Entry restrictions, industry evolution, and dynamic efficiency: Evidence from commercial banking, Research paper no. 9630, Federal Reserve Bank of New York.

Kaparakis, E.I., S.M. Miller, and A.G. Noulas, 1994, Short-run cost inefficiency of commercial banks: A flexible stochastic frontier approach, Journal of Money, Credit, and Banking 26, 875-893.

Kwan S.H. and R.A. Eisenbeis, 1995, An analysis of inefficiency in banking: A stochastic cost frontier approach, Working paper, Working Papers in Applied Economics, Federal Reserve Bank of San Francisco.

McAllister, P.H. and D. McManus, 1993, Resolving the scale efficiency puzzle in banking, Journal of Banking and Finance 17, 389-405.

Mester, L.J., 1992, Traditional and nontraditional banking: An information-theoretic approach, Journal of Banking and Finance 16, 545-566.

Mester, L.J., 1993, Efficiency in the savings and loan industry, Journal of Banking and Finance 17, 267-286.

Mester, L.J., 1994, How efficient are Third District banks? Business Review, Federal Reserve Bank of Philadelphia, 3-18.

Mester, L.J., 1996, A study of bank efficiency taking into account risk-preferences, Journal of Banking and Finance 20, 1025-1045.

Mitchell, K. and N.M. Onvural, 1996, Economies of scale and scope at large commercial banks: Evidence from the Fourier flexible functional form, Journal of Money, Credit, and Banking 28, 178-199.

Pi, L. and S.G. Timme, 1993, Corporate control and bank efficiency, Journal of Banking and Finance 17, 515-530.

Sealey, C. and J. Lindley, 1977, Inputs, outputs, and a theory of production and cost at depository financial institutions, Journal of Finance 32, 1251-1266.

Shaffer, S., 1994, A revenue-restricted cost study of 100 large banks, Applied Financial Economics 4, 193-205.

Shaffer, S. and E. David, 1991, Economies of superscale in commercial banking, Applied Economics 23, 283-293.

Spong, K., R. Sullivan, and R. DeYoung, 1995, What makes a bank efficient? A look at financial characteristics and bank management and ownership structure, Financial and Industry Perspectives, Federal Reserve Bank of Kansas City.

Stevenson, R.E., 1980, Likelihood functions for generalized stochastic frontier estimation, Journal of Econometrics 13, 57-66.

Table 1

Variables Employed in the Cost, Standard Profit, and Alternative Profit Functions
Means and Standard Deviations for 1995 only

(All financial variables measured in 1000's of constant 1994 dollars,
Prices of financial assets and liabilities are measured as interest rates.)

Symbol	Definition	Mean	Std Dev
	Dependent Variables		
C	Variable operating plus interest costs, includes costs of purchased funds, deposits, and labor	13,466	105,671
π	Variable profits, includes revenues from loans and securities less variable costs	8,628	66,767
	Variable Input Prices		
w_1	Price of purchased funds (jumbo CDs, foreign deposits, federal funds purchased, all other liabilities except core deposits).	.0410	.0111
w_2	Price of core deposits (domestic transactions accounts, time and savings)	.0284	.0081
w_3	Price of labor (1000's of constant dollars per employee)	32.5	6.8
	Variable Output Quantities (Cost and Alternative Profit Functions Only)		
y_1	Consumer loans (installment and credit card and related plans)	38,179	298,130
y_2	Business loans (all other loans)	164,952	1,489,552
y_3	Securities (all non-loan financial assets, i.e., Gross Total Assets $- y_1 - y_2 - z_2$)	114,916	838,231
	Variable Output Prices (Standard Profit Function Only)		
p_1	Price of consumer loans	.0926	.0329
p_2	Price of business loans	.0898	.0126
p_3	Price of securities	.0468	.0087

(Table 1, p. 2)

Symbol	Definition	Mean	Std Dev
	Fixed Netput Quantities		
z_1	Off-balance-sheet guarantees (commitments, letters of credit, etc.) measured using Basle Accord risk weights to be risk-equivalent to loans.	26,367	445,427
z_2	Physical capital (premises and other fixed assets)	4,818	38,909
z_3	Financial equity capital	26,686	184,880
	Environmental Variables		
NPL	Nonperforming loans (past due at least 90 days or on nonaccrual basis) divided by total loans	.0258	.0217
STNPL	Weighted average of NPL for the state, using proportions of the loans issued by banks in the state as the weights.	.0220	.0043
Num. observations, cost and alternative profit regressions, 1995		9,002	
Num. observations, standard profit regressions, 1995		8,378	

Notes: All stock values are real quantities as of the December call report and all prices are flows over the year divided by these stocks. Because the price data are subject to error from this procedure, we eliminate observations in which the prices on assets and liabilities (which are interest rates) are more than 2.5 standard deviations from the mean value for that year. Similarly, we eliminate observations in which liability and asset rates are more than .10 above and more than .50 above the one-year Treasury rate, respectively. The standard profit function uses fewer observations because these procedures eliminated some output price data. We also eliminated observations in which equity was below 1% of gross total assets because the data for such banks are suspicious. From these regressions, efficiency is reported only for the 5,949 observations in which all of these data plus data on other variables in Table 2 are available for every year 1990-1995. All of the continuous variables that can take on the value 0 have 1 added before taking logs in specifying the cost and profit regressions. This applies to the y's, z's, NPL, and STNPL. For π, an additional adjustment was made because profits can take on negative values (see text).

Table 2

Measured Bank X-Efficiency Using Various Econometric Models U.S. Banks 1990-1995

(weighted mean efficiencies, standard deviations in parentheses, correlations)

Model Specification	Cost Efficiency	Standard Profit Efficiency	Alternative Profit Efficiency
Preferred Model: Distribution-Free Approach, Fourier-Flexible Specification, Includes State Average Nonperforming Loan Ratios and Equity Capital.	.868 (.062)	.549 (.208)	.463 (.195)
Same as Preferred Model, Except Uses Translog Specification In Place of Fourier-Flexible Specification.	.860 (.063)	.539 (.207)	.452 (.194)
Rank-Order Correlation with Preferred Model:	.979	.995	.995
Same as Preferred Model, Except Includes Bank's Own NPL Ratio Instead of State Average STNPL.	.866 (.062)	.550 (.209)	.469 (.198)
Rank-Order Correlation with Preferred Model:	.992	.999	.997
Same as Preferred Model, Except Includes Bank's Own NPL Ratio in Addition to State Average STNPL.	.866 (.062)	.550 (.208)	.471 (.197)
Rank-Order Correlation with Preferred Model:	.992	.999	.999
Same as Preferred Model, Except Excludes both STNPL and NPL.	.869 (.061)	.546 (.207)	.461 (.196)
Rank-Order Correlation with Preferred Model:	.999	.9999	.998
Same as Preferred Model, Except Excludes Equity Capital.	.869 (.062)	.088 (.287)	.106 (.301)
Rank-Order Correlation with Preferred Model:	.834	.215	.425
Same as Preferred Model, Except Stochastic Frontier Approach, Half-Normal Distribution.[a]	.942 (.023)	— —	.351[b] (.709)
Rank-Order Correlation with Preferred Model:	.988	—	.912
Number of Observations:	5949	5949	5949

[a]For the stochastic frontier approach to yield meaningful efficiency estimates, the cost or profit residuals must have the correct skew, (rightward for cost, leftward for profit). In some cases this did not occur, including standard profit efficiency in all six years, cost efficiency in 1992, and alternative profit efficiency in 1994 and 1995. The stochastic frontier cost efficiency estimate is therefore based on data from 1990, 1991, 1993, 1994 and 1995 only, and alternative profit efficiency is based on data from 1990, 1991, 1992 and 1993.
To be consistent with the distribution free approach, the stochastic frontier efficiencies were calculated by comparing firms to the best observed expected value of $ln u$ so that the best-practice firm has an efficiency of 1.

[b]Three banks were eliminated from this calculation because they had negative potential profits, and the profit efficiency ratios are meaningless in these cases. This leaves a sample size of 5946.

Table 3

Cost Scale and X-Efficiency Estimates by Size Class (Weighted Means)

Bank Size (GTA)	# of Banks	Scale Efficiency	% Below Efficient Scale	t*	X-Efficiency	Total Cost Efficiency	C/GTA
0-$50M	2218	.856	94.3%	2.200	.851	.728	.0482
$50M-$100M	1794	.842	96.5%	2.363	.870	.733	.0473
$100M-$300M	1344	.818	97.3%	2.523	.873	.715	.0466
$300M-$1B	392	.786	99.8%	2.815	.876	.688	.0453
$1B-$10B	171	.782	99.8%	2.986	.860	.671	.0436
> $10B	30	.782	94.4%	2.673	.872	.680	.0427
Total	5949	.795	97.3%	2.723	.868	.689	.0443

Notes: All figures (except C/GTA, which is weighted by GTA) are weighted by each bank's predicted costs; consequently the efficiencies indicate the proportion of the sample banks' costs that are used efficiently, with respect to scale efficiency, frontier efficiency and total cost efficiency, the product of scale and frontier efficiency.

The percent below efficient scale refers to the percentage of banking costs that are incurred by firms below efficient scale.

t^* is the bank's ratio of the cost-efficient size to its actual size, so that $t^* > 1$ indicates unexploited scale economies, and $t^* < 1$ indicates diseconomies of scale.

C/GTA is average cost or total cost divided by gross total assets, is included as a raw-data version of cost efficiency. It has an unweighted mean of .0473 and a standard deviation of .0050.

Because the Fourier-flexible form has numerous local minima and maxima, the scale-efficient point for each bank's product mix was discovered by means of grid search over the interval [0.5, 5.0]. We limit the interval in order to avoid extrapolating too far from the observed data, but this interval should allow for any realistic short-term expansion or contraction prospects. Fortunately the limits on the interval were generally non-binding, as only 3 banks out of 5949 had t^* values at one of the endpoints (see Berger, Leusner, and Mingo, 1996 for more details on this method).

Table 4

Correlations Among the Efficiency Measures and Raw-Data Measures of Performance†

	COST EFFICIENCY	STANDARD PROFIT EFFICIENCY	ALTERNATIVE PROFIT EFFICIENCY	C/GTA	ROA
STANDARD PROFIT EFFICIENCY	0.019 (0.145)				
ALTERNATIVE PROFIT EFFICIENCY	-0.167 (0.000)	0.794 (0.000)			
C/GTA	-0.206 (0.000)	-0.119 (0.000)	-0.235 (0.000)		
ROA	0.247 (0.000)	0.122 (0.000)	0.177 (0.000)	-0.279 (0.000)	
ROE	0.205 (0.000)	0.469 (0.000)	0.334 (0.000)	-0.220 (0.000)	0.726 (0.000)

C/GTA is average cost, i.e., total cost divided by gross total assets, a raw-data version of cost efficiency. It has an unweighted mean of 0.047 and a standard deviation of 0.005 over 1990-1995.

ROA is return on assets, i.e., net income divided by gross total assets. It has a mean of 0.011 and a standard deviation of 0.004 over 1990-1995.

ROE is return on equity, i.e., net income divided by equity. It has a mean of 0.016 and a standard deviation of 0.047 over 1990-1995.

†Spearman correlation coefficients, with p-values of the tests for zero correlation in parentheses.

Table 5

Cost, Standard Profit, and Alternative Profit X-Inefficiencies Relative to Potential Profits,[a] Gross Total Assets, and Equity

		Inefficiency / Potential Profits			Inefficiency / GTA			Inefficiency / Equity		
Bank Size (GTA)	# of Banks	Cost[b]	Standard Profit[c]	Alternative Profit[c]	Cost[b]	Standard Profit[c]	Alternative Profit[c]	Cost[b]	Standard Profit[c]	Alternative Profit[c]
0-$50M	2218	.165	.414	.412	.007	.018	.018	.074	.185	.184
		(.085)	(.157)	(.217)	(.003)	(.011)	(.011)	(.039)	(.071)	(.090)
$50M-$100M	1794	.143	.382	.362	.006	.016	.015	.065	.173	.164
		(.074)	(.140)	(.188)	(.003)	(.009)	(.009)	(.034)	(.065)	(.078)
$100M-$300M	1344	.182	.472	.732	.006	.015	.023	.066	.172	.267
		(.092)	(.220)	(.276)	(.003)	(.008)	(.010)	(.034)	(.072)	(.102)
$300M-$1B	392	.184	.449	.775	.006	.013	.023	.068	.167	.288
		(.107)	(.229)	(.311)	(.003)	(.008)	(.009)	(.039)	(.089)	(.101)
$1B-$10B	171	.204	.456	.694	.006	.013	.020	.082	.182	.278
		(.113)	(.224)	(.276)	(.003)	(.008)	(.009)	(.045)	(.092)	(.104)
> $10B	30	.184	.468	.628	.005	.014	.018	.078	.199	.267
		(.135)	(.200)	(.248)	(.003)	(.005)	(.006)	(.051)	(.073)	(.084)
Total	5949	.185	.451	.639	.006	.014	.020	.075	.183	.259
		(.114)	(.208)	(.303)	(.003)	(.008)	(.009)	(.044)	(.081)	(.102)

[a]The potential profits denominator in the first set of comparisons is calculated using the standard profit function. Using the alternative profit function yields similar results.
[b]Cost inefficiency is predicted cost minus minimum cost.
[c]Profit efficiency is potential profits minus predicted profits.
Note: All figures are weighted averages of the values for all banks in the sample. If an individual bank's value is given by x/z, then the figures in the table are $\sum x_i / \sum z_i$.
Numbers in parentheses are standard deviations.

Table 6

Variables Employed as Potential Correlates of Efficiency
(One observation per bank averaged over 1990-1995 unless otherwise indicated)
(All financial variables measured in 1000's of constant 1994 dollars)

Symbol	Definition	Mean	Std Dev
	Bank Size Variables		
SMLBANK	Dummy, equals one if bank has GTA below $100 million. Excluded from the regressions as the base case.	.673	.449
MEDBANK	Dummy, equals one if bank has GTA of $100 million to $1 billion.	.293	.432
LARBANK	Dummy, equals one if bank has GTA of $1 billion to to $10 billion.	.028	.157
HUGBANK	Dummy, equals one if bank has GTA over $10 billion. over $10 billion.	.005	.067
	Organizational Form/Governance		
MERGED	Dummy, equals one if bank survived one or more bank-level mergers during the period (i.e., absorbed the assets of one or more other banks).	.132	.338
ACQUIRED	Dummy, equals one if bank was acquired by a new high holder bank holding company during the period.	.108	.310
INBHC	Dummy, equals one if bank is owned by a bank holding company.	.761	.407
MUL_LAY	Dummy, equals one if the bank is in a multiple-layered BHC, i.e., the direct holder is not the high holder.	.087	.251
OUTST	Dummy variable, equals one if bank's high holder is located in another state.	.055	.211
PUB_TRADED	Dummy, equals one if the bank's high holder is registered with the SEC for public trading.	.228	.407
INSIDE	Proportion of stock owned by board members and their relatives. Reported only for banks with over 50% of the banking assets of publicly traded organizations (126 observations).	.159	.187
OUTSIDE	Proportion of stock owned by outside owners with share blocks greater than 5%. Same restrictions as **INSIDE** (126 observations).	.058	.083

(Table 6, p. 2)

Symbol	Definition	Mean	Std Dev
	Other Bank Characteristics		
AGE	Number of years the bank existed before 1990.	67.6	33.6
LOAN/GTA	Loans divided by gross total assets (GTA).	.532	.130
SUB_DER	Dummy, equals one if the total notional value of the bank's swaps, forwards, futures, and similar contracts exceeds 5% of GTA.	.025	.155
PF/GTA	Purchased funds (deposits > $100,000, foreign deposits, federal funds purchased, subordinated debt, other non-deposit liabilities) to GTA ratio.	.119	.061
SDROA	Standard deviation over time of the bank's annual return on assets.	.003	.003
SDROE	Standard deviation over time of the bank's annual return on equity (used only in robustness checks to substitute for SDROA).	.036	.043
	Market Characteristics		
HERF	Herfindahl index of local market concentration.	.248	.155
SHARE	Bank's share of local market deposits (used only in robustness checks to substitute for HERF).	.175	.197
INMSA	Dummy, equals one if the bank is in a Metropolitan Statistical Area.	.361	.472
STGROW	Real state income growth (decimal).	.007	.002
STUNEMP	State unemployment rate (decimal) (used only in robustness checks to substitute for STGROW).	.065	.013
	State Geographic Restrictions on Competition		
UNITB	Dummy, equals one for unit banking states (the six-year average will be at most .167, there are no UNIT banking states after 1990).	.010	.040
LIMITB	Dummy, equals one for limited branching states.	.473	.468
STATEB	Dummy, equals one for statewide branching states. Excluded from the multiple regressions as the base case.	.516	.462

(Table 6, p. 3)

Symbol	Definition	Mean	Std Dev
	State Geographic Restrictions (con't.)		
LIMITBHC	Dummy, equals one for states with limits on expansions of multibank holding companies. As of 1990, all states permitted some multibank holding company activity, so the excluded case is that the state allows statewide holding company powers.	.518	.500
NOINTST	Dummy, equals one for states that do not allow interstate expansions of multibank holding companies.	.032	.095
ACCESS	Proportion of nation's banking assets in states that are allowed to enter the state (equals proportion of national assets in the state for states that do not allow interstate banking).	.503	.276
	Primary Federal Regulator		
FED	Dummy, equals one if the bank's primary federal regulator is the Federal Reserve.	.085	.268
FDIC	Dummy, equals one if the bank's primary federal regulator is the FDIC.	.608	.481
OCC	Dummy, equals one if the bank's primary federal regulator is the OCC. Excluded from the multiple regressions as the base case.	.307	.455
	Raw Data Measures of Performance[1]		
C/GTA	Total cost divided by gross total assets, a raw-data version of cost efficiency.	.047	.005
ROA	Return on assets: ratio of net income to gross total assets.	.011	.004
ROE	Return on equity: ratio of net income to equity.	.116	.047
Num. Observations		5,949	

[1]The raw data measures of performance are included in the analysis as alternative measures of efficiency as robustness checks of the more complicated frontier efficiency estimates. They are included in a separate correlation analysis but are excluded from the regression analysis as being completely endogenous.

Table 7
Regression Analysis of the Potential Correlates of Efficiency†

Dep Var	COST EFFICIENCY (1)	(2)	STANDARD PROFIT EFFICIENCY (3)	(4)	ALTERNATIVE PROFIT EFFICIENCY (5)	(6)
Bank Size Variables						
SMLBANK	0.883**	-0.021**	0.184**	0.083**	0.375**	0.206**
	(137.355)	(-12.682)	(10.929)	(15.799)	(19.890)	(38.908)
MEDBANK	0.902**	0.021**	0.036*	-0.083**	0.117**	-0.200**
	(132.626)	(12.118)	(1.994)	(-15.099)	(5.885)	(-35.778)
LARBANK	0.899**	0.0123*	0.009	-0.038*	0.077**	-0.147**
	(97.130)	(2.536)	(0.358)	(-2.444)	(2.859)	(-8.708)
HUGBANK	0.908**	0.009	-0.040	-0.089*	0.093*	-0.132**
	(60.940)	(0.773)	(-1.028)	(-2.492)	(2.133)	(-3.349)
Organizational Form/Governance						
MERGED	-0.004	0.007**	0.023**	0.004	0.031**	-0.052**
	(-1.532)	(3.168)	(3.606)	(0.511)	(4.303)	(-6.631)
ACQUIRED	0.006*	0.010**	-0.002	0.019*	0.007	0.013
	(2.223)	(3.989)	(-0.335)	(2.387)	(1.005)	(1.508)
INBHC	0.004*	0.012**	0.055**	0.079**	0.057**	0.039**
	(2.062)	(6.318)	(10.549)	(13.609)	(9.909)	(5.974)
MUL_LAY	0.011**	0.020**	0.040**	0.032**	0.0314**	-0.017
	(3.303)	(6.709)	(4.525)	(3.314)	(3.201)	(-1.561)
OUTST	0.010*	0.021**	-0.006	0.015	-0.011	-0.046**
	(2.562)	(5.996)	(-0.529)	(1.273)	(-0.960)	(-3.687)
PUB_TRADED	0.008**	0.017**	0.025**	0.032**	0.013*	-0.038**
	(3.691)	(9.064)	(4.433)	(5.448)	(2.114)	(-5.825)
Other Bank Characteristics						
AGE	0.00003	0.0001**	-0.0006**	-0.0008**	-0.0009**	-0.001**
	(0.900)	(4.283)	(-.9.776)	(-10.734)	(-12.373)	(-13.256)
LOAN/GTA	-0.019**	0.005**	0.755**	0.656**	0.426**	0.260**
	(-2.943)	(0.867)	(45.731)	(39.678)	(23.059)	(12.860)
SUB_DER	-0.019**	0.004	0.015	-0.013	0.053**	-0.084**
	(-3.128)	(0.771)	(0.930)	(-0.856)	(2.924)	(-4.917)
PF/GTA	0.024	0.032**	-0.143**	-0.131**	-0.224**	-0.404**
	(1.771)	(2.620)	(-3.970)	(-3.335)	(-5.558)	(-9.396)
SDROA	-2.517**	-2.905**	-2.89**	0.258	-3.876**	1.718
	(-9.120)	(-10.788)	(-3.995)	(0.299)	(-4.796)	(1.805)
Market Characteristics						
HERF	-0.015**	-0.018**	0.023	0.035**	0.067**	0.130**
	(-2.610)	(-3.614)	(1.527)	(2.243)	(4.039)	(7.572)
INMSA	-0.007**	0.001	0.003	-0.008	0.021**	-0.042**
	(-3.368)	(0.926)	(0.491)	(-1.474)	(3.665)	(-7.405)
STGROW	-1.610**	−1.640**	4.172**	1.843	6.038**	8.967**
	(-3.220)	(3.602)	(3.185)	(1.274)	(4.125)	(5.632)

†Columns (1), (3), and (5) report multivariate regression coefficients with t-statistics in parentheses. Note that the OCC and STATEB dummy variables and a constant term are omitted from the multivariate regressions to avoid perfect collinearity. Columns (2), (4), and (6) report univariate regression coefficients, where each regression included the variable and an intercept term, with t-statistics in parentheses.

*Significant at the 5% level. **Significant at the 1% level.

(Table 7†, p. 2)

Dep Var	COST EFFICIENCY (1)	(2)	STANDARD PROFIT EFFICIENCY (3)	(4)	ALTERNATIVE PROFIT EFFICIENCY (5)	(6)
State Geographic Restrictions on Competition						
UNITB	-0.058**	-0.059**	0.337**	0.450**	0.426**	0.603**
	(-2.768)	(-3.106)	(6.115)	(7.547)	(6.922)	(9.196)
LIMITB	0.008**	0.005**	-0.013**	-0.020**	-0.041**	-0.006
	(4.450)	(3.146)	(-2.641)	(-3.912)	(-7.449)	(-1.131)
STATEB	------	0.005**	-----	0.017**	-----	0.002
		(-2.911)		(3.300)		(0.349)
LIMITBHC	-0.0008	-0.004*	0.032**	0.007	0.032**	0.030**
	(-0.462)	(-2.380)	(6.737)	(1.442)	(6.064)	(5.730)
NOINTST	-0.046**	-0.036**	-0.028	-0.089**	-0.039	-0.047
	(-4.745)	(-4.497)	(-1.053)	(-3.536)	(-1.374)	(-1.682)
ACCESS	-0.011**	0.0009	0.054**	-0.027**	0.079**	-0.006
	(-3.090)	(0.315)	(6.008)	(-3.080)	(7.891)	(-0.593)
Primary Federal Regulator						
FED	0.008*	0.008**	-0.003	0.002	-0.013	-0.015
	(2.559)	(2.947)	(-0.412)	(0.252)	(-1.433)	(-1.536)
FDIC	-0.0006	-0.007**	-0.008	0.018**	-0.019**	0.031
	(-0.340)	(-4.193)	(-1.784)	(3.514)	(-3.808)	(5.544)
OCC	-----	0.004**	------	0.020**	------	0.029**
		(2.696)		(-3.863)		(-4.953)
Adj R-sq‡	0.067		0.364		0.348	
No. Banks	5949	5949	5949	5949	5949	5949

†Columns (1), (3), and (5) report multivariate regression coefficients with t-statistics in parentheses. Note that the OCC and STATEB dummy variables and a constant term are omitted from the multivariate regressions to avoid perfect collinearity. Columns (2), (4), and (6) report univariate regression coefficients, where each regression included the variable and an intercept term, with t-statistics in parentheses.

‡The adjusted R-squared is computed as $1-[(N-1)/(N-K)](1-R^2)$ where N=number of observations, K=number of parameters, R^2= sum of squared errors from the regression/sum of squared deviations of the dependent variable from its mean. This adjusted R-squared is equivalent to that obtained from the model in which SMLBANK is replaced with an intercept term.

*Significant at the 5% level. **Significant at the 1% level.