FEDERAL RESERVE BANK OF PHILADELPHIA

Ten Independence Mall
Philadelphia, Pennsylvania 19106-1574
(215) 574-6428, www.phil.frb.org

Working Papers

Research Department

WORKING PAPER NO. 97-5

EFFICIENCY AND PRODUCTIVITY CHANGE IN THE U.S. COMMERCIAL BANKING INDUSTRY: A COMPARISON OF THE 1980s AND 1990s

Allen N. Berger
Board of Governors of the Federal Reserve System
and
The Wharton Financial Institutions Center, University of Pennsylvania

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

May 1997

WORKING PAPER 97-5

EFFICIENCY AND PRODUCTIVITY CHANGE IN THE U.S. COMMERCIAL BANKING INDUSTRY: A COMPARISON OF THE 1980s AND 1990s

Allen N. Berger
Board of Governors of the Federal Reserve System
and
The Wharton Financial Institutions Center, University of Pennsylvania

and

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

May 1997

Presented at the Conference on Service Sector Productivity and the Productivity Paradox,
Centre for the Study of Living Standards, Ottawa, Ontario, April 11-12, 1997

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 our discussant Jack Triplett and participants at the Conference on Service Sector Productivity and the Productivity Paradox sponsored by the Centre for the Study of Living Standards for helpful comments, Dave Humphrey for helpful comments and advice, Mike Mohr, and Bob Yuskavage for helpful advice, Seth Bonime and Chris Malloy for excellent research assistance, and Sally Burke for expert editorial 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: aberger@frb.gov. To Mester at Research Department, Federal Reserve Bank of Philadelphia, Ten Independence Mall, Philadelphia, PA 19106-1574; phone: (215) 574-3807; fax: (215) 574-4364; email: Loretta.Mester@PHIL.frb.org.

Efficiency and Productivity Change in the U.S. Commercial Banking Industry: A Comparison of the 1980s and 1990s

Abstract

We investigate efficiency and productivity growth of the U.S. banking industry over the latter part of the 1980s and first part of the 1990s using comprehensive data on U.S. commercial banks. Cost efficiency decreased slightly between the 1980s and 1990s and large banks showed a sizable decline in profit efficiency. Total predicted production costs increased over both the 1980s and 1990s, reflecting cost productivity declines. Changes in business conditions led to cost declines over both periods. Total predicted profits increased in the 1980s and 1990s, with the entire change reflecting increased profit productivity. Changing business conditions led to small declines in profits.

Efficiency and Productivity Change in the U.S. Commercial Banking Industry:
A Comparison of the 1980s and 1990s

I. INTRODUCTION

This paper investigates the efficiency and productivity change of the U.S. banking industry over the latter part of the 1980s and first part of the 1990s. We are motivated by the importance of the banking sector to the economy as a whole and by the tremendous transformation that this industry is undergoing. Our other goal is to promote a set of concepts and techniques for analyzing efficiency and productivity change that is applicable to any industry in any country—cost, standard profit, and alternative profit efficiency and productivity change. We believe that the joint application of these methods—each of which provides some independent information and each of which is valid under somewhat different assumptions—provides a comprehensive look at industry efficiency and productivity growth that would be difficult to match with any single cost or profit function method or with any single cost, profit, or productivity ratio.

The banking industry accounts for 3% of GDP according to the National Income and Product Accounts, but this industry's efficiency and productivity growth has far greater than proportional import to the functioning of the economy as whole. Banks provide liquidity, payments, and safekeeping services for depositors and intermediate their funds into investment and working capital resources for nonfinancial industries. Banks play a special role in funding small businesses, which often have very limited access to other sources of external finance. Banks also maintain a smoothly functioning payments system that allows financial and real resources to flow relatively freely to their highest-return uses. Thus, inefficiency or productivity problems in the banking industry can have significant consequences across a wide range of nonfinancial firms and industries.

The most striking aspect of the transformation of the U.S. banking industry is its consolidation. The number of banks has declined from more than 14,000 in 1980 to about 9500 in 1996, mostly from mergers and acquisitions. Over the 1980s and first half of the 1990s, there were over 3,500 mergers in which two or more banks were consolidated under a single charter, as well as more than 5,800 acquisitions in which banks retained their individual charters but became owned by different bank holding companies (BHCs). In the first half of the 1990s, bank mergers involved about 20% of industry assets each year, and BHCs that acquired other banks constituted about another 20% of the industry in each of these years (Berger, Saunders, Scalise, and Udell, 1997). This consolidation is primarily a result of the lifting of geographic barriers to competition, and this trend will likely continue under the Riegle-Neal Interstate Banking and Branching Efficiency Act, which allows virtually nationwide branching as of June 1, 1997. This relaxation of barriers to competition and the resulting increase in competitive pressures are expected to drive institutions toward becoming more efficient and to raise productivity growth in the long term. Other changes in the banking industry include the expansion of banking powers, changes in capital requirements, the removal of interest rate ceilings on deposits, innovations in financial engineering, increased competition from nonbank rivals and foreign competitors, and the application of new information-processing technologies. These fundamental changes likely have important consequences for industry efficiency and productivity change and, by extension, important consequences for the economy as whole, but quantification of these effects has been fairly elusive.

There has been a substantial research effort into studying bank efficiency, but a more modest effort into measuring productivity change in this important industry. The efficiency and productivity growth studies have differed in the efficiency/productivity concepts used, measurement methods used, and potential correlates specified that may explain some of the variations in efficiency and productivity change. Since the studies varied in so many dimensions, it is difficult to determine the important factors leading to the differences in results. To try to resolve this problem for efficiency measurement, Berger and Mester (hereafter BM) (1997) studied the sources of differences in bank efficiency by applying different efficiency concepts, different measurement methods, and different potential correlates of efficiency to the same data set. Their findings suggested that each of three efficiency concepts—cost, standard profit, and alternative profit efficiency—adds some independent informational value. The efficiency results were remarkably robust to the different measurement techniques, different functional forms, and various treatments of output quality used. Treatment of equity capital was found to be an important consideration.

In this paper, we extend the analysis to compare bank efficiency and productivity change between the last six years of the 1980s and the first six years of the 1990s. Our data set includes annual information from 1984 through 1995 on virtually all U.S. commercial banks—over 145,000 annual bank observations in all.

Cost, standard profit, and alternative profit efficiency and productivity change are measured. Cost efficiency uses a cost function to measure how close a bank's cost is to what a best-practice bank's cost would be for producing the same output quantities when facing the same market prices of variable inputs and other variables. Similarly, cost productivity change measures the proportional change in costs from one year to another for the same variables, all measured at the industry mean. Standard profit efficiency and productivity change are based on a variable profit function that includes revenues as well as costs, and takes as given market prices for outputs, rather than treating the output quantities as given. The standard profit measures embody everything that is in the cost measures plus the revenue and cost effects of failing to choose output mix and scale optimally. Alternative profit efficiency/productivity change keeps the revenues as in the standard profit approach, but specifies output quantities instead of output prices. As discussed in BM (1997), the alternative profit approach may be informative when some of the assumptions underlying the cost and standard profit measures are not met.

Our efficiency/productivity measures are based on economic optimization in response to relative prices, rather than optimization based just on the physical technology. The cost, alternative profit, and standard profit functions account for input prices and the standard profit function accounts for output prices. So measures based on these functions take account of allocative inefficiency as well as technical inefficiency. Thus, a bank's use of more of a relatively high priced input than is optimal will be measured as an inefficiency or as a decline in productivity, even if no more output could be produced from the same inputs. Efficiency/productivity measures that do not account for relative prices and allocative inefficiency are incomplete in an important way in our opinion.

The total change in industry cost or profits over time may be decomposed into two elements—(1) productivity growth, which is the proportional change in cost or profit for a given set of "business conditions," and (2) the change in costs or profits due to the "business conditions" themselves, which are simply the exogenous variables specified in the cost or profit function. Both elements are measured below. Productivity growth may be further decomposed into two parts—(1) the movement of the best-practice efficient frontier, and (2) the change in the average degree of efficiency or the dispersion of firms away from this frontier. The efficient frontier is measured in terms of what the best-practice firms do, rather than any "true" minimum costs or maximum profits that cannot be measured. The movements of the frontier may be driven by technological change, such as improvements in information-processing technologies or improvements in applied finance that allow banks to make better investments at lower cost. The frontier may also be moved by regulatory changes that affect costs or profits, such as the deregulation of interest rates or relaxation of geographic entry barriers. The location of the best-practice frontier also depends on competitive conditions, since even the managers of the best-practice banks may reduce effort or pursue goals other than cost minimization or profit maximization if competition is lax. In our analysis, we measure total productivity change by using average-practice cost and profit functions that are based on data from all banks, and we measure movement of the best-practice efficient frontier by estimating cost and profit functions for only the best-practice quarter of the banks.

It is quite possible for the efficient frontier to either improve or worsen over time, and prior literature has found movements in both directions. Although technological retrogression is unlikely, changes in regulation and competitive conditions can result in worse performance of the best-practice banks, and this has sometimes been found in the literature. Changes in regulation and competition can affect the average industry inefficiency or dispersion from the frontier as well, which may also affect measured productivity change positively or negatively.

Our findings show some decrease in cost efficiency between the 1980s and 1990s. Considering all banks, there is little change in profit efficiency between the 1980s and 1990s, but large banks did show a sizable decline. The total predicted cost of production increased over the 1980s and over the 1990s, reflecting a decline in cost productivity. Changes in business conditions led to cost declines over the two periods. Total predicted profits increased in the 1980s and 1990s, with the entire change reflecting increased profit productivity. Changing business conditions led to small declines in profits.

Section II reviews prior information on bank productivity change. Section III lays out our efficiency and productivity concepts, and Section IV gives the design of our empirical analysis. Section V shows our efficiency results, and Sections VI and VII display our productivity results for the whole sample and for separate size classes, respectively. Section VIII draws conclusions.

II. PRIOR ANALYSIS OF BANK PRODUCTIVITY CHANGE

The literature on efficiency in the financial services industry has been extensively reviewed elsewhere (Berger and Humphrey, 1997), so we omit a review here. Less is known about productivity growth in the banking industry. We first briefly discuss what some government agencies do, and then cover a handful of academic studies to give a flavor of the research, but this is by no means a comprehensive review.

Government agencies typically measure productivity growth by the change in a ratio of an output index to an input index. The BEA measures gross product originating (GPO) by industry (often referred to as "value-added" by industry), divided by labor hours in that industry. For banking, the value-added is determined primarily as -1 times the value of interest paid on deposits less the interest received on loans, which is reckoned to be the value of services provided free to depositors. Productivity growth between benchmarking dates (every 5 years) is measured by the change in this value-added per labor hour. However, for non-benchmark years, GPO in banking is estimated by simply extrapolating the labor input measure, so no productivity growth can be estimated. The BLS measures physical banking output using a "number of transactions" approach based on time deposits (number of savings accounts, club accounts, CDs, and other deposits), demand deposits (number of checks written and cleared, and number of electronic funds transfers), loans (number of new loans extended), and trust accounts (number of these accounts), each weighted by the proportion of man hours used in each activity. Again, labor hours are used as the denominator of the productivity index. The use of labor hours in both the BEA and BLS measures does not account for the increase in "outsourcing" of certain back-office operations to holding company affiliates, which are still measured in bank costs, but not as bank labor. See Griliches (1992), Dean and Kunze (1992), Mohr (1992), and Yuskavage (1996) for more information on these measures. Fixler and Zieschang (1997) propose a promising method for measuring bank output that incorporates both the financial approach of the BEA and physical production approach of the BLS; Fixler and Hancock (1997) propose a method of measuring the credit services of banks.

These ratios do not have some of the benefits of the cost and profit productivity measures in our opinion. The use of a single index of bank output does not fully account for allocative inefficiencies or differences in product mix across banks that affect banks' costs and profits. For example, a product mix that generates higher revenue will not be measured as an increase in productivity growth, as it would in a standard profit productivity measure. Because the output measures are not regressed on prices or other control variables, the government productivity indexes do not separate out "business conditions," but rather measure productivity change and changes in business conditions together. Similarly, the government measures do not separate out movements of the best-practice efficient frontier from changes in the inefficiency or dispersion from the frontier, so it is difficult to draw conclusions about whether technology is improving versus whether more banks are taking advantage of existing technology. Finally, there may be difficulties with using labor as the single input. It has been shown that the ratio of labor to costs has changed dramatically over time (Berger and Humphrey, 1992), and that BHCs have moved many of their operations outside the bank itself, so that costs are incurred, but the labor may be employed elsewhere in the holding company (Berger, Kashyap, and Scalise 1995).

Several academic studies have measured cost productivity. Our discussion will slightly reinterpret some of their results using our own terminology. The literature often calls shifts in the best-practice frontier "technological" change, but we prefer to keep explicit the distinction between technology used by the best-practice banks and the theoretically best technology available.

Berger and Humphrey (1992) used the thick frontier approach to compare bank cost efficiency and to study movements in the best-practice frontier between 1980, 1984, and 1988 using a complete sample of U.S. banks. These three years correspond to pre-, mid-, and post-deregulation of the deposit side of banking. As discussed above, shifts in the frontier over time may reflect changes in the technology used by the best-practice banks, regulation, or other competitive conditions. The authors found that when the shifts are not adjusted for changes in business conditions, average costs increased for all but the very largest efficient banks in the 1980-84 interval, followed by decreases in average costs for all sizes in the 1984-88 period. The increase in costs in the earlier period was larger for the smaller banks in the sample. This reflected the increase in deposit rates that occurred in the 1980-84 period because of deregulation, and the fact that smaller banks use a larger proportion of deposits to fund their assets than do larger banks. The decline in average costs in the later interval reflected a general decline in market rates that affected deposit rates and rates on purchased funds about equally. When the shifts in the average cost frontiers were adjusted for changes in the exogenous variables in the cost function and market rates, an increase is still shown for the 1980-84 period, but a decrease is no longer shown for the 1984-88 period.

Bauer, Berger, and Humphrey (hereafter, BBH) (1993) used a panel data set of 683 banks with over $100 million in assets from states that allowed branching and that were continuously in existence during 1977-88 to estimate total factor cost productivity growth for the best-practice banks. Their estimates ranged from an average annual growth rate over the period of −2.28% to 0.16%, depending on the estimation method used. The poor productivity growth was attributed to higher costs of funding due to high market rates, elimination of deposit rate ceilings, and increased competition from nonbank financial intermediaries, which increased demand for funds and reduced the supply of deposits. Instead of compensating for this increase in input price by closing branches or substituting ATMs, increased competition forced banks to provide more in the way of convenience. Hence, banks increased the number of branches over the 1980s, in addition to paying higher deposit rates and providing the ATM innovation. The increase in deposit rates, increase in nonbank competition, and better convenience all made consumers of bank services better off, but because quality of service is difficult to account for in the estimation, the higher quality showed up as a decrease in productivity.

Humphrey (1993) used the same data set to investigate the effect on costs from shifts in the cost function from 1977-88. Measures were derived in three ways: from a simple time trend; from a time-specific index; and from annual shifts in cross-section cost functions. All three methods yielded similar estimates, with shifts in the cost function implying cost increases averaging 0.8% to 1.4% per year, and small banks (with assets of $100-$200 million) experiencing a larger decline than large banks. Humphrey attributed the decline in productivity to deregulation of deposit rates. As the cost of deposits rose, banks found themselves to be "overbranched," and the productivity of their deposit base declined. As support for this hypothesis, he found that in the pre-deregulation period (1977-80), productivity increased, while during deregulation (1981-82), productivity declined substantially, and in the post-deregulation period (1983-88), it showed little change.

Again using the same data set, Humphrey and Pulley (1997) estimated changes in predicted bank profits in the 1977-88 period and decomposed the changes that occurred after deregulation (1984-88) into internal, bank-initiated adjustments to the new regulatory structure and external, contemporaneous changes in banks' business conditions. They found that for banks with assets over $500 million, the rise in profits from the 1977-81 period to the 1981-84 period resulted from a shift in the profit function, by changing deposit and loan prices, and changing the use of labor, capital, and funding inputs in response to deregulation. In contrast, shifts in the profit function did not account for the rise in larger banks' profits from 1981-84 to 1985-88. Rather, this rise was due to higher levels of deposits, loans, other assets, and changes in input prices. For smaller banks, with assets between $100 million and $500 million, the authors found that there was little increase in profits between 1977-81 and 1981-84, and in the later period, their experience was similar to that of larger banks. Repeating their analysis abstracting from changes in bank efficiency, they found larger changes in profits, but the pattern was the same, with shifts in the profit frontier dominating the earlier period and changes in business conditions dominating the latter.

Devaney and Weber (1996) investigated whether the market structure of rural banking markets affected the banks' productivity growth over 1990-93. They used linear programming, rather than stochastic techniques, to calculate the Malmquist productivity index, which decomposes productivity changes into changes in efficiency, shifts in the production function, and changes in the scale of operations. Since these measures are based on quantities of outputs and inputs without regard to prices, there is no way to determine whether banks became more or less productive in an economic sense or responded more or less appropriately to market price signals. The authors found positive productivity growth at rural banks over 1990-93. Shifts in the production frontier were the driving force of this productivity growth.

Wheelock and Wilson (1996) also used the linear programming approach to investigate bank productivity growth, decomposing the change in productivity into its change in efficiency and frontier shift components. They found that larger banks (assets over $300 million) experienced productivity growth between 1984-93, while smaller banks experienced a decline. Average inefficiency remained high in the industry, since banks were not able to adapt quickly to changes in technology, regulations, and competitive conditions.

III. THE EFFICIENCY AND PRODUCTIVITY CONCEPTS

We focus on three economic concepts from which efficiency and productivity measures will be derived—cost, standard profit, and alternative profit. We prefer these concepts, since they are based on economic optimization in reaction to market prices, rather than being based solely on the use of technology.[1]

1. Cost. The cost function relates variable costs to the prices of variable inputs, the quantities of variable outputs and any fixed netputs, environmental factors, and random error, as well as efficiency:

$$\ln C = f_C(w,y,z,v) + \ln u_C + \ln \varepsilon_C . \tag{1}$$

C measures variable costs, f_C is some functional form, 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, v is a set of environmental variables that may affect performance, u_C denotes an inefficiency factor that may raise costs above the best-practice level, and ε_C denotes random error. The term, $\ln u_C + \ln \varepsilon_C$, is treated as a composite error term, and the various efficiency measurement techniques differ in how they distinguish the inefficiency term, $\ln u_C$, from the random error term, $\ln \varepsilon_C$.

We define the cost efficiency of bank b as the estimated cost needed to produce bank b's output vector

---

[1]See BM (1997) for a fuller discussion.

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.,

$$\text{Cost EFF}^{b} = \frac{\hat{C}^{min}}{\hat{C}^{b}} = \frac{\exp[\hat{f}_C(w^b, y^b, z^b, v^b)] \times \exp[\ln \hat{u}_C^{min}]}{\exp[\hat{f}_C(w^b, y^b, z^b, v^b)] \times \exp[\ln \hat{u}_C^{b}]} = \frac{\hat{u}_C^{min}}{\hat{u}_C^{b}}, \tag{2}$$

where $\hat{u}_C^{min}$ is the minimum $\hat{u}_C^{b}$ across all banks in the sample. Cost efficiency ranges over (0,1], and equals one for a best-practice firm within the observed data.

The total change over time in an industry's costs reflects changes in productivity and changes in the exogenous variables that affect costs, which we refer to as "business conditions." The change in predicted total costs (with random error removed) between periods t and t+1 is given by:

$$\text{Total Change}_{Cost} = \frac{\hat{C}_{t+1}(X_{t+1}, u_C)}{\hat{C}_t(X_t, u_C)} = \frac{\hat{C}_{t+1}(X_t, u_C)}{\hat{C}_t(X_t, u_C)} \times \frac{\hat{C}_{t+1}(X_{t+1}, u_C)}{\hat{C}_{t+1}(X_t, u_C)}, \tag{3}$$

where $X_t \equiv (w_t, y_t, z_t, v_t)$, u_C is the inefficiency factor, and $\hat{C}$ is the estimated "average-practice" cost function, which is estimated using all banks and embodies inefficiency. Note that this is a gross cost change, so, e.g., if costs go up by 5%, our total cost change would be 1.05.

The first term on the right-hand-side of (3) represents productivity change, i.e.,

$$\text{Productivity Change}_{Cost} = \frac{\hat{C}_{t+1}(X_t, u_C)}{\hat{C}_t(X_t, u_C)}, \tag{4}$$

and the second term on the right-hand-side of (3) represents changes in business conditions, X:

$$\text{Business Conditions Change}_{Cost} = \frac{\hat{C}_{t+1}(X_{t+1}, u_C)}{\hat{C}_{t+1}(X_t, u_C)}. \tag{5}$$

A change in the scale of operations would be included as a change in business conditions for the cost function, since output quantities are included in the exogenous variables of the cost function.

As noted above, productivity change comprises the movement of the best-practice frontier and change in inefficiency, or dispersion from the frontier. Movement in the best-practice cost frontier is given by:

$$\text{Frontier Change}_{Cost} = \frac{\exp[\hat{f}_{Ct+1}(X_t)]}{\exp[\hat{f}_{Ct}(X_t)]}, \tag{6}$$

where $\hat{f}_C$ is the estimated best-practice cost frontier, which excludes the inefficiency term.

In this paper, we will always estimate the total change, productivity change, and frontier change evaluated at the same X, the mean values of $\ln w_t$, $\ln y_t$, $\ln z_t$, and $\ln v_t$ over all the banks in the sample for the year in question. It is important to remember that frontier change here means changes in the practices used by the best-practice firms, which depend on technology, regulation, and competitive conditions. As discussed above, any of the changes we measure can be positive or negative.

2. Standard profit. The standard profit function takes variable output prices as given but allows output quantities to vary, so that it accounts for revenues that can be earned by varying outputs as well as inputs:

$$ln\,(\pi + \theta) = f_{\pi}(w,p,z,v) + ln\,u_{\pi} + ln\,\varepsilon_{\pi}\,. \tag{7}$$

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; $ln\,\varepsilon_{\pi}$ represents random error; and $ln\,u_{\pi}$ represents inefficiency that reduces profits.

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 variables). Efficiency measures based on the profit function allow for inefficiencies in the choice of inputs (like the cost function does) but also in the choice of outputs when responding to output prices or to any other arguments of the profit function. Thus, profit efficiency is more comprehensive than cost efficiency and gives a better measure of the overall performance of the firm.

Standard profit efficiency is the ratio of the predicted actual profits to the predicted maximum profits of a best-practice bank facing the same business conditions, net of random error, or the proportion of maximum profits that are actually earned:

$$\text{Std}\ \pi\ \text{EFF}^{\,b} = \frac{\hat{\pi}^{b}}{\hat{\pi}^{max}} = \frac{\left\{\exp[\hat{f}_{\pi}(w^{\,b},p^{\,b},z^{\,b},v^{\,b})] \times \exp[\,ln\,\hat{u}_{\pi}^{\,b}]\right\} - \theta}{\left\{\exp[\hat{f}_{\pi}(w^{\,b},p^{\,b},z^{\,b},v^{\,b})] \times \exp[\,ln\,\hat{u}_{\pi}^{\,max}]\right\} - \theta} \tag{8}$$

where $\hat{u}_{\pi}^{max}$ is the maximum value of u_{π}^{b} in the sample. Profit efficiency ranges over $(-\infty,1]$, and equals one for a best-practice firm within the observed data. Unlike cost efficiency, profit efficiency can be negative, since firms can throw away more than 100% of their potential profits.

We define the total change, productivity change, business conditions change, and frontier change for standard profits in a similar way to costs. For example, standard profit frontier change is the ratio of predicted best-practice profits for year t+1 to those for year t, both evaluated at the mean values of $ln\,w_t$, $ln\,p_t$, $ln\,z_t$, and $ln\,v_r$.

3. Alternative profit. The alternative profit function is identical to the standard profit function in (7) except that y replaces p in the function f, yielding different values for the inefficiency and random error terms:

$$ln\,(\pi + \theta) = f_{a\pi}(w,y,z,v) + ln\,u_{a\pi} + ln\,\varepsilon_{a\pi}\,. \tag{9}$$

The efficiency and change statistics are calculated in the same way as the standard profit measures except for this change in the arguments of the profit function.

The concept of alternative profit may be helpful when some of the assumptions underlying the cost and standard profit functions are not met.[2]

IV. EMPIRICAL DESIGN

Our data set includes annual information on virtually all U.S. commercial banks from 1984 through 1995; thus, we have over 145,000 observations (over 80,000 observations for the 1980s and over 65,000 for the 1990s). Most of the data are from the Reports of Income and Condition (Call Reports).[3]

1. Variables included in the cost and profit functions. Table 1 gives the definitions of all the variables specified in the cost and alternative profit functions, as well as their sample means and standard deviations for the periods 1984-89 and 1990-95. The variable input prices, w, include the interest rates on purchased funds and core deposits as well as the price of labor. These prices are calculated as exogenous market-average prices that a bank faces. In each MSA or non-MSA county in which a bank operates, we constructed the weighted average of the bank prices in the market excluding the bank's own price, where the weights are each bank's share of the total input usage in the market and each bank's price is its expenditures on the input divided by its input quantity. The bank's input price is then the weighted average of the prices faced in each of its markets, where the weight is the bank's share of its deposits in that market. The variable outputs, y, include consumer loans, business loans, real estate 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. The prices of these outputs, which appear in the standard profit function, are constructed as market-average prices, similar to the way in which input prices are constructed.[4]

We specify off-balance-sheet items, physical capital, and financial equity capital as fixed netputs, z. For

---

[2]It would be useful, e.g., if (i) there are substantial unmeasured differences in the quality of banking services; (ii) outputs are not completely variable; (iii) output markets are not perfectly competitive; and/or (iv) output prices are not accurately measured. See BM (1997) for further details.

[3]Major Call Report changes implemented in 1984 made it difficult to include the earlier part of the 1980s.

[4]We use the state-average input or output price if the market-average price is likely to be, or appears to be, unreliable (e.g., if the bank accounts for most of a market's deposits, the market-average price would not be representative of the market, since it is constructed by omitting the bank). See the footnote to Table 1.

the off-balance-sheet items, we use the Basle Accord risk weights on the assumption that costs and revenues should be approximately proportional to the perceived credit risk on which these weights are based. We specify these items as fixed instead of variable primarily because of the difficulty of obtaining accurate price information. Since physical capital is slow to adjust, we treat it as a fixed input as well.

We include equity capital because it directly affects costs, as an alternative source of funding for bank operations. It may have an indirect effect via the risk premium a bank pays for uninsured deposits and nondeposit funds, since financial capital provides a cushion against insolvency.[5] The specification of equity may also reduce scale bias, as discussed in BM (1997).

Among the environmental variables, v, we include the log of the market-average nonperforming to total loan ratio, *ln* MNPL, and its square, ½ $(ln \text{ MNPL})^2$, since dealing with exogenous loan problems raises costs and lower profitability. To control more directly for risk, we also include SDROA, the standard deviation of a bank's annual return on assets over the five, four, or three years prior to the date in question (choosing the longest for which data were available) and NOSDROA, which equals one if there were fewer than three years of data for the bank and zero otherwise.[6] We also specify controls for the geographic restrictions on bank competition including branching restrictions (unit banking (UNITB), limited branching (LIMITB), with statewide branching as the base case); the degree of in-state holding company expansion permitted (LIMITBHC); whether out-of-state holding company expansion is prohibited (NOINTST); and the proportion of the banking industry's assets held in states allowed to enter the bank's own state (ACCESS, ½ACCESS^2). Other controls include the Herfindahl index of local deposit market concentration (HERF); state income growth (STINC, ½STINC^2); whether the bank is located in a metropolitan area (INMSA), and the identity of a bank's federal regulator (FED, FDIC, with OCC as the base case).

2. Functional form. The results presented below are based on the translog functional form.[7] Our

---

[5]Hughes and Mester (1993 and forthcoming) and Hughes, et al. (1995) discuss the importance of including equity in cost and profit function models.

[6]If fewer than three years of data were available, then SDROA was set equal to zero.

[7]BM (1997) found that both the translog and the Fourier-flexible functional form, which is a global approximation that includes a standard translog plus Fourier trigonometric terms, yielded essentially the same average level and dispersion of measured efficiency, and both ranked the individual banks in almost the same order.

specification of the cost function is:

$$\begin{aligned} ln\,(C/w_3z_3) = \alpha &+ \sum_{i=1}^{2} \beta_i\, ln\,(w_i/w_3) + \tfrac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} \beta_{ij}\, ln\,(w_i/w_3)\, ln\,(w_j/w_3) + \sum_{k=1}^{4} \gamma_k\, ln\,(y_k/z_3) \\ &+ \tfrac{1}{2}\sum_{k=1}^{4}\sum_{m=1}^{4} \gamma_{km}\, ln\,(y_k/z_3)\, ln\,(y_m/z_3) + \sum_{r=1}^{2} \delta_r\, ln\,(z_r/z_3) + \tfrac{1}{2}\sum_{r=1}^{2}\sum_{s=1}^{2} \delta_{rs}\, ln\,(z_r/z_3)\, ln\,(z_s/z_3) \\ &+ \sum_{i=1}^{2}\sum_{k=1}^{4} \eta_{ik}\, ln\,(w_i/w_3)\, ln\,(y_k/z_3) + \sum_{i=1}^{2}\sum_{r=1}^{2} \rho_{ir}\, ln\,(w_i/w_3)\, ln\,(z_r/z_3) \\ &+ \sum_{k=1}^{4}\sum_{r=1}^{2} \tau_{kr}\, ln\,(y_k/z_3)\, ln\,(z_r/z_3) + \sum_{i=1}^{17} \xi_i v_i + ln\, u_C + ln\, \varepsilon_C \end{aligned} \qquad (10)$$

where we have suppressed time subscripts; (y_k/z_3), (z_r/z_3), and the MNPL variables (included in v) have 1 added for every firm in order to avoid taking the natural log of zero; and the standard symmetry restrictions apply.[8]

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 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, since the exogenous variables are identical to those for the cost function. For the standard profit function, the terms containing the variable output quantities, $ln\,(y_k/z_3)$, are replaced by the corresponding output prices, $ln\,(p_k/w_3)$.

All of the cost, profit, input price, and output price terms are normalized by the last input price, w_3, in order to impose linear homogeneity on the model.[9] We also specify all of the cost, profit, output quantities, and fixed netput quantities as ratios to the fixed equity input, z_3, to control for heteroskedasticity, to help control for scale biases in the estimation, and to give the models more economic interpretation, since the dependent variable in the profit functions becomes approximately ROE.

3. Technique for estimating efficiency. We estimate bank efficiency for our two subperiods, 1984-89

---

[8]We exclude factor share equations, which embody restrictions imposed by Shephard's Lemma or Hotelling's Lemma, because these would impose the undesirable assumption of no allocative inefficiencies.

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

and 1990-95, using the distribution-free method (Berger, 1993). This method uses less arbitrary assumptions than the more popular stochastic frontier approach to disentangle inefficiencies from random error, and it gives a better idea of a bank's longer-term performance, since the method bases efficiency estimates on several years of data. The method assumes there is a core inefficiency or average inefficiency for each firm over time. The core inefficiency is distinguished from random error by assuming that core inefficiency is persistent over time, while random errors tend to average out over time. Our efficiency estimates are based on six years of data. One banking study (DeYoung, 1997) found that this length of time balanced out concerns that the period be long enough so that random errors average out and that the period be short enough so that core inefficiency was constant over the period.

The cost, standard profit, and alternative profit equations are estimated separately for each year, 1984-95, allowing the coefficients to vary to reflect changes in technology, regulation, and market environment. For each subperiod (1984-89 and 1990-95), the average residual for each bank b in the subperiod is formed, which is an estimate of $ln\ u_C^b$, $ln\ u_\pi^b$, or $ln\ u_{a\pi}^b$ for the subperiod, depending on the equation.[10] The resulting estimates of the inefficiency terms, $ln\ \hat{u}_C^b$, $ln\ \hat{u}_\pi^b$, and $ln\ \hat{u}_{a\pi}^b$, for the two subperiods, along with their minimum or maximum values $ln\ \hat{u}_C^{min}$, $ln\ \hat{u}_\pi^{max}$, and $ln\ \hat{u}_{a\pi}^{max}$ in the subperiod, are then used to obtain each bank's efficiency ratio for the subperiod.[11] We also compute efficiency measures for small banks (assets averaging under $1 billion in 1994 dollars over the six-year subperiod in question) and large banks (all other banks), separately, in two different ways. One is based on the frontier estimated using the full sample. The other is based on separately estimated

---

[10]Despite the assumption that random error for each bank averages out to zero over time, 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 for each subperiod. 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 the 5th or 95th percentile, respectively. Truncation is performed within each of ten size classes (by gross total assets) to reduce the effects of persistently good or bad luck for these banks relative to firms of their size.

[11]Because the costs and profits in the dependent variables are expressed in terms of the ratios to w_3z_3, the $\exp[\hat{f}(\cdot)]$ terms in the efficiency ratios are replaced by $w_3z_3 \times \exp[\hat{f}(\cdot)]$ for cost efficiency and $w_3z_3 \times \exp[\hat{f}(\cdot)] - \theta$ for profit efficiency, where $\hat{f}(\cdot)$ is the predicted part of the cost or profit function. To offset the nonlinearities introduced by exponentiating and including the θ terms, in all but one case 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. In one case (the standard profit function for 1990), this adjustment is not made, since the profit numbers are close to zero and the multiplicative adjustment distorted the efficiency ratio.

frontiers for small and large banks, which allow for the possibility that small and large banks may be using different production technologies or are producing slightly different products. For example, the typical $1 million of commercial loans for a small bank may represent services to 10 customers, whereas for a large bank it might represent one-tenth of one customer.

4. Technique for estimating productivity changes and related measures. To estimate the total changes in costs or profits, productivity changes, and changes in business conditions for each year, we estimate "average-practice" cost and profit functions without respect to inefficiency, and then apply the formulas in equations (3), (4), and (5) for the cost function, and comparable ones for standard and alternative profit. We present annualized measures for 1984-95, 1984-89, 1989-95, and for each pair of consecutive years between 1984 and 1995. Because different sized banks might have exhibited differences in productivity growth, we repeat the above analysis for small banks and large banks separately (using their own frontiers). (Here, small banks are those with less than $1 billion (1994 dollars) in total assets in the year in question and large banks are all others.)

We also measure the frontier changes, as in equation (6), by reestimating the cost and profit function models using only the best-practice efficient banks. By using only the best banks to estimate the parameters, we are able to trace shifts in the best-practice efficient frontier over time, abstracting from changes in inefficiency or dispersion from the frontier.

We use a version of the thick frontier method to choose which banks are on the frontier each year (Berger and Humphrey, 1991). We divide the banks based on their residuals from estimating the average-practice cost and profit functions, which use the data on all banks operating in the year. Banks with residuals in the "best" 25% in their size category (i.e., lowest cost or highest profit) are assumed to be best-practice efficient for that year (we use 10 asset size categories). We then reestimate the best-practice cost and profit functions for these most efficient quarter of banks (assuming that the residuals represent random error and not differences in inefficiency). These estimated functions are treated as the cost and profit frontiers.

V. EFFICIENCY RESULTS

Table 2 shows the means and standard deviations of the efficiencies estimated for all banks and for large and small banks for 1984-89 (hereafter, the 1980s) and 1990-95 (hereafter, the 1990s).[12] The mean cost

---

[12]The cost and profit function coefficient estimates are available from the authors upon request.

efficiency for banks of all sizes in the 1980s is 0.806, which suggests that about 19.4% of costs were wasted, on average, relative to a best-practice firm. The 0.806 figure is consistent with the literature, which typically finds about 80% cost efficiency for the 1980s. For the 1990-95 period, we estimate average cost efficiency decreased very slightly to 0.770, i.e., 77%.[13] The decrease in cost efficiency between the 1980s and 1990s might be attributed to banks having to readjust after the large numbers of mergers many were involved in at the end of the 1980s and early 1990s. However, the efficiency estimates give a measure of the dispersion of the banks from the best-practice frontiers for the two subperiods and these frontiers may be different. The lower efficiency estimate for the 1990s need not indicate that banks are using more resources to produce a given level of output than they were in the 1980s, since the cost frontier may have improved. Our estimates of frontier change address this issue. Also note that this slight decrease in efficiency may not be statistically significant.

Looking at the estimates for small and large banks separately (but still using the same frontier), we find that the difference between the efficiency estimates for large and small banks was 1.2 percentage points in the 1980s and 2.5 percentage points in the 1990s. The lower efficiency of smaller banks might be due to the fact that these banks tend to operate in more concentrated markets and are more insulated from competitive pressures that can drive firms toward efficiency. Or it might reflect the fact that small banks are producing slightly different products from large banks or are using a different technology. The estimates of small and large bank efficiencies based on their own frontiers can be used to check this possibility. In general, the estimates are fairly similar to those obtained using the frontier estimated for all banks. As can be seen, relative to their own frontiers, small banks showed higher average efficiency, i.e., were less dispersed from their own frontier, than were large banks in the 1980s, but they had nearly identical efficiency measures in the 1990s.

The mean efficiencies for the standard and alternative profit functions for the 1980s and 1990s for all banks 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. The point estimate for standard profit efficiency is the same in the 1980s and 1990s, while there is only a small increase for alternative profit efficiency. Our estimates are well within the observed range from the few other profit efficiency studies. The standard deviations of the profit

---

[13] This estimate is lower than the one found in BM (1997), which used a similar model specification but a less comprehensive set of banks, used bank-specific rather than market-average prices, and omitted many of our control variables in order to investigate potential correlates with efficiency.

efficiencies are about 23 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 to 8 percentage points. Although the average profit efficiencies appear to be lower than the average cost efficiency, these ratios are not directly comparable because they are reported in terms of different denominators (predicted actual costs versus potential profits).[14]

There do appear to be differences between small and large banks. Although in the 1980s, small and large banks have similar profit efficiencies, in the 1990s, small banks are considerably more profit efficient than the large banks, with efficiency estimates (based on their own frontiers) 11 to 19 percentage points higher. Since small and large banks have similar cost efficiencies (based on their own frontiers) in the 1990s, the difference in the profit efficiencies must reflect differences in revenue efficiencies. Hence, large banks appear to be less efficient either in setting output prices or output levels than smaller banks are. This may reflect differences in the types of customers large banks serve or the types of products offered. Looking at changes in efficiency over time, we see that for the large banks, efficiency decreased between the 1980s and 1990s: standard profit efficiency for large banks, measured relative to their own frontier, fell from about 50% in the 1980s to less than 40% in the 1990s; similarly, alternative profit efficiency for large banks declined from 60% in the 1990s to less than 50% in the 1990s. For small banks, the change in efficiency between the 1980s and 1990s is quite minor.

We also examined 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 (available from the authors). The general picture is the same for the 1980s and 1990s. Standard and alternative profit efficiency are highly positively and statistically significantly correlated with each other ($\rho = 0.914$ in the 1980s and 0.929 in the 1990s), as expected. The correlation of cost efficiency and the two profit efficiencies is statistically significant but at a much lower level, in the 0.2 range for the 1980s and in the 0.15 range for the 1990s. 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 $\equiv$ total cost/gross total assets, and positively and significantly correlated with the standard profitability ratios, ROA and

[14]BM (1997) recasted their estimates in terms of similar denominators and found that profit inefficiency was larger than cost efficiency, as expected (i.e., there are revenue inefficiencies).

ROE. These findings suggest that our efficiency measures are robust and are not simply the consequences of our specifications or methods.

VI. RESULTS ON PRODUCTIVITY CHANGES AND RELATED MEASURES

Table 3 reports the total predicted change in costs and profits and the decompositions into productivity changes and changes in business conditions. These average-practice statistics are based on cost and profit functions estimated using all banks and allow for inefficiency (i.e., the dispersion of banks from the frontier) to change over time, as well as for the frontier to shift. The frontier change statistics also reported in Table 3 are based on estimations using only banks in the best cost and profit quarters; they reflect shifts in the frontier only and abstract from changes in efficiency. The first row in each table shows changes for the entire period 1984-95 (at an annual rate), with 1984 serving as base year. The next two rows show the changes for the 1980s and 1990s, respectively (both at an annual rate). The other rows in the tables give the annual changes. For the total change, each subperiod change is the geometric mean of the yearly changes in that subperiod. To abstract from the volatility of the measures, we will focus on the changes for the 1980s and 1990s. The annual changes are very volatile from year to year, and probably less meaningful. This is especially true for the change in profits. For example, 1989 was a good year for bank profitability but 1990 was a particularly bad year, and this is shown in the very large drop in profit productivity between 1989 and 1990, and equally very large increase between 1990 and 1991. Realized average profit was actually negative in 1990, but predicted profit for the average-practice banks evaluated at the means of the exogenous variables was a small positive number, resulting in a nine-fold increase in profit between 1990 and 1991. Because predicted profit was so low in 1990, it was relatively easy for banks to earn nine times the amount the next year.[15] We expect the change in profit and profit productivity to be more volatile than the change in costs and cost productivity, since profit is more volatile from year to year than is cost.

As can be seen in Table 3, the total predicted cost of producing banks' output was little changed over the entire period. Cost rose in the 1980s by about 1% annually and then fell about the same amount in the 1990s. Looking at the decomposition, we see that changes in productivity and changes in business conditions worked

---

[15]For the best-practice banks, predicted standard profit was negative for 1989; hence, we do not compute the total change or frontier change for 1989-90.

in opposite directions. Cost productivity decreased in both subperiods, with the 1980s showing a 6% annual increase in costs and the 1990s showing a 9.5% annual increase. Business conditions for bank costs improved in both the 1980s and 1990s, but more so in the 1990s—in the 1980s, exogenous changes reduced costs by about 5%, while in the 1990s, these factors reduced costs by about 10%. This reflects in part the drop in interest rates in the 1990s.

Looking at the frontier change column, we see that shifts in the frontier increased bank costs in the 1980s and more so in the 1990s, and that the effect of frontier changes on costs are generally less than the productivity changes that allow for dispersion from the frontier. In other words, banks on the frontier shifted toward more costly production on average, and at the same time, the dispersion from this frontier increased. This result is consistent with what we reported in Table 2: that the average degree of efficiency decreased slightly between the 1980s and 1990s. Both the shift in frontier technology and the decrease in efficiency worked to reduce productivity between the 1980s and 1990s.

While these decreases in cost productivity and frontier change might seem surprising, our results for the 1980s are consistent with those of Berger and Humphrey (1992), BBH (1993), and Humphrey (1993). These papers attribute the poor and sometimes negative productivity change of banks in the 1980s to deregulation of deposit rates and other increases in competition. BBH argue that as banks lost some of their market power over depositors, their costs rose. Instead of being able to offset these costs with lower operating costs, e.g., by closing branches and replacing them with ATMs, the increased competition the banks faced for deposits caused them to increase the convenience offered to depositors in the form of a larger number of branches. This increase in convenience would not appear as increased cost productivity, but may show up as increased profit productivity if banks were able to capture some of the benefit. Otherwise, bank customers may have taken all of it themselves. We look at this momentarily.

What is surprising in our cost results is that there is negative cost productivity and frontier change even in the 1990s, when one would have thought the effect of deregulation of deposit rates would have already been incorporated. There are several possible explanations: (i) competitive pressures on banks have continued to increase and banks have continued to react by increasing output quality, which raises costs, and banks have not yet seen a revenue benefit from higher quality; (ii) increased competition might also have changed the banks' clientele to one that is more costly to serve; (iii) new technology has changed the way banks do business and the fixed costs of this new technology have not yet been recovered; (iv) the mergers many of these banks have experienced might also have caused disruptions in production, temporarily raising costs (even if in the long run the mergers will lead to reduced costs);[16] and (v) differences in the productivity of small and large banks and changes in the proportions of these banks in the industry over time might be affecting the average productivity measures reported for all banks. We examine some of these differences in the next subsection.

Before turning to our profit results, it might be useful to compare our cost productivity results with the BLS measures of commercial bank (SIC code 602) productivity. One might expect that our measures would agree, but the BLS measures indicate productivity increased an average of 2.49% per year in the 1984-89 period and 3.48% per year in the 1989-95 period. As explained in Section II above, the BLS statistics do not adequately capture differences in product mix across banks, do not account for allocative inefficiencies, do not separate out business conditions or shifts in the frontier, and focus on labor productivity rather than total factor productivity.

Turning to profits, the change in predicted standard profit and in predicted alternative profit are quite similar over the entire period 1984-95, with standard profit showing an annual increase over 11% and alternative profit an increase of almost 7%. Comparing the change in profit estimates with our estimates for cost shows that while costs increased over the period, revenues increased even more, so that profitability increased. This suggests that banks were receiving some payment for the increased convenience they were offering their customers.

The decomposition shows that the increase in profits came entirely from increased productivity. Changing business conditions led to a small decline in profits in the 1980s and 1990s of about 3% to 4% annually each subperiod. Note that since, on average for all sized banks, there was little change in the level of profit efficiency between the 1980s and 1990s, the change in profit productivity mainly reflects an upward shift in the profit frontier. In terms of the standard profit function, frontier changes reflect changes in the banks' choice of output levels and use of inputs, while changes in business conditions reflect changes in output and input prices, and other exogenous variables. In contrast, for the alternative profit function, frontier change reflects changes

[16] To test this possibility, we reestimated our measures using a sample of banks that were not involved in mergers for the year in question and at least the three previous years. Our results are little changed, suggesting that mergers are not a likely explanation.

in output prices and the use of inputs, while changes in business conditions reflect changes in output levels, input prices, and other exogenous variables. Our results differ from those of Humphrey and Pulley (1997), who did not find positive productivity change between 1981-84 and 1985-88.

For both profits, but more so for standard profit, the increase in the 1980s was larger than the increase in the 1990s. This probably reflects the fact that 1989, the boundary year for our two subperiods, was a very good year for bank profits. This would tend to inflate the profit change for the 1984-89 period and deflate the change for the 1989-95 period. Had we instead chosen 1990 as the boundary, the 1980s would look much worse, and the 1990s much better.

VII. PRODUCTIVITY CHANGES AND RELATED MEASURES: LARGE VS. SMALL BANKS

We next examine whether small and large banks had different experiences in terms of productivity and related changes over the 1980s and 1990s. The estimates reported in Table 4 are based on functions and frontiers estimated for small and large banks separately. We do find some differences between small and large banks, the main one being that the differences between the 1980s and 1990s are more pronounced for the large banks than for the small banks (some of this might be because the sample of large banks is small), even though on average over the entire period from 1984-95 small and large banks experienced similar changes in costs and profits.

There is little difference in the total change in costs for small and large banks over the 1980s and 1990s subperiods. Cost productivity declined for both small and large banks in both subperiods, but the decline was considerably larger for large banks in the 1990s, with large banks experiencing about a 19% increase in costs due to frontier change in the 1990s. A countervailing effect was that changing business conditions were more beneficial to large banks than to small banks in terms of their costs. The results for standard profits are similar to those for alternative profits, with both small and large banks showing an increase of about 6-7% in standard profits in the 1980s and 1990s. In terms of standard profits, productivity change, and especially frontier change, was higher for larger banks in the 1990s than 1980s. In contrast, productivity change for small banks was relatively stable for the 1980s and 1990s.

VIII. CONCLUSIONS

This paper investigates the efficiency and productivity growth of the U.S. banking industry for the latter part of the 1980s and the first part of the 1990s. Our results shed some light on how banks are responding to changes in technology, regulations, and business conditions, an important question given the tremendous transformation this industry is experiencing and the importance of the banking sector to the economy as a whole. We base our analysis on the cost, standard profit, and alternative profit concepts, which we believe are superior to other measures (like simple ratios of cost to input or output to input), since they derive from economic optimization in response to relative prices rather than optimization based just on the physical technology.

We find that cost efficiency for banks of all sizes averaged about 80% over 1984-89 and declined slightly to 77% over 1990-95 (this decrease may not be statistically significant). Profit efficiency estimates are similar for the standard and alternative profit function with banks showing about 50% efficiency in both subperiods. In the 1980s, small and large banks have similar profit efficiencies, but in the 1990s large banks are considerably less profit efficient than small banks. Although the average profit efficiencies appear to be lower than the average cost efficiency, these ratios are not directly comparable because they are reported in terms of different denominators (predicted actual costs versus potential profits).

Banks experienced an adverse shift of the cost frontier toward higher costs and negative cost productivity change in both the 1980s and 1990s. Changes in business conditions provided a countervailing effect, leading to lower costs. So, on net, there was little change in total costs. The profit story is different, with banks experiencing beneficial shifts in the profit frontier and profit productivity growth over the 1980s and 1990s. Changes in business conditions hurt the banks in terms of profits. The differences between the cost and profit results point to the importance of examining all three measures—cost, standard profit, and alternative profit. The profit measures give a more comprehensive view of efficiency and productivity change in an industry than the cost measures, since they incorporate revenues and the cost effects of failing to choose output mix and scale optimally.

REFERENCES

Bauer, P.W., A.N. Berger, and D.B. Humphrey (1993) ‘Efficiency and productivity growth in U.S. banking,’ in *The Measurement of Productive Efficiency: Techniques and Applications*, ed. H.O. Fried, C.A.K. Lovell, and S.S. Schmidt (Oxford: Oxford University Press) 386-413

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. and R. DeYoung (1997) “Problem loans and cost efficiency in commercial banks,” *Journal of Banking and Finance* 21

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 *Measurement Issues In the Service Sectors*, ed. Z. Griliches (Chicago: National Bureau of Economic Research, University of Chicago Press) 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., A.K Kashyap, and J.M. Scalise (1995) ‘The transformation of the U.S. banking industry: What a long, strange trip it’s been,’ *Brookings Papers on Economic Activity* (2), 55-218.

Berger, A.N., A. Saunders, J.M. Scalise, and G.F. Udell (1997) ‘The effects of bank mergers and acquisitions on small business lending,’ Working Paper, Federal Reserve Board, Washington, DC

Berger, A.N. and L.J. Mester (1997) ‘Inside the black box: What explains differences in the efficiencies of financial institutions?’ *Journal of Banking and Finance* 21

Dean, E.R., and K. Kunze (1992) ‘Productivity measurement in service industries,’ in *Measurement Issues In the Service Sectors*, ed. Z. Griliches (Chicago: National Bureau of Economic Research, University of Chicago Press) 73-101

Devaney, M. and W. Weber (1996) ‘Productivity growth, market structure, and technological change: Evidence from the rural banking sector,’ mimeo, Southeast Missouri University

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*

Fixler, D. and D. Hancock (1997) 'Measuring bank credit services: Large versus small U.S. banks,' mimeo, Federal Reserve System

Fixler, D. and K. Zieschang (1997) 'The productivity of the banking sector: Integrating financial and production approaches to measuring financial service output,' mimeo, U.S. Bureau of Labor Statistics

Griliches, Z. (1992) 'Introduction,' in *Measurement Issues In the Service Sectors*, ed. Z. Griliches (Chicago: National Bureau of Economic Research, University of Chicago Press) 1-22.

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. 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 (forthcoming) 'Bank capitalization and cost: Evidence of scale economies in risk management and signaling,' *The Review of Economics and Statistics*

Humphrey, D.B. (1993) 'Cost and technical change: Effects from bank deregulation,' *Journal of Productivity Analysis* 4, 9-34

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

Mohr, M.F. (1992) 'Recent and planned improvements in the measurement and deflation of services outputs and inputs in BEA's Gross Product Originating estimates,' in *Measurement Issues In the Service Sectors*, ed. Z. Griliches (Chicago: National Bureau of Economic Research, University of Chicago Press) 25-68

Wheelock, D.C. and P.W. Wilson (1996) 'Technical progress, inefficiency, and productivity change in U.S. banking, 1984-1993,' mimeo, Federal Reserve Bank of St. Louis

Yuskavage, R.E. (1996) 'Improved estimates of gross product by industry, 1959-94,' *Survey of Current Business* 76, 133-155.

TABLE 1

Variables Employed in the Cost, Standard Profit, and Alternative Profit Functions
Means and Standard Deviations for 1984-89 and 1990-95

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

		1984-89 (80,014 obs.)		1990-95 (65,560 obs.)	
Symbol	**Definition**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
	Dependent Variables				
C	Variable operating plus interest costs, includes costs of purchased funds, deposits, and labor.	18,439	236,847	17,118	224,197
π	Variable profits, includes revenues from loans and securities less variable costs.	3615	36,853	3835	53,901
	Variable Input Prices				
w_1	Market-average price of purchased funds (jumbo CDs, foreign deposits, federal funds purchased, all other liabilities except core deposits).	0.0623	0.0155	0.0446	0.0157
w_2	Market-average price of core deposits (domestic transactions accounts, time and savings).	0.0533	0.0148	0.0319	0.0129
w_3	Market-average price of labor (1000's of constant dollars per employee).	31.6	5.45	33.2	6.24
	Variable Output Quantities (Cost and Alternative Profit Functions Only)				
y_1	Consumer loans (loans to individuals).	29,973	233,319	34549	270,195
y_2	Business loans (all loans other than consumer and real estate loans).	86,519	1,310,505	84,762	1,333,755
y_3	Real estate loans.	49,881	412,305	82,457	652,836
y_4	Securities (all non-loan financial assets, i.e., Gross Total Assets $- y_1 - y_2 - y_3 - z_3$).	105,593	1,072,634	138,480	1,423,613
	Variable Output Prices (Standard Profit Function Only)				
p_1	Market-average price of consumer loans.	0.102	0.0273	0.0971	0.0241
p_2	Market-average price of business loans.	0.125	0.0420	0.106	0.0359
p_3	Market-average price of real estate loans.	0.0899	0.0174	0.0799	0.0145
p_4	Market-average price of securities.	0.0616	0.0117	0.0368	0.0167

TABLE 1, CON'T.

Symbol	Definition	1984-89 (80,014 obs.) Mean	Std. Dev.	1990-95 (65,560 obs.) Mean	Std. Dev.
	Fixed Netput Quantities				
z_1	Off-balance-sheet items (commitments, letters of credit, derivatives, etc.) measured using Basle Accord risk weights to be risk-equivalent to loans.	28,802	620,588	39,429	847,100
z_2	Physical capital (premises and other fixed assets).	4004	38,063	5102	52,371
z_3	Financial equity capital.	16,790	138,960	25,196	212,944
	Environmental Variables				
MNPL	Market-average of nonperforming loans (past due at least 90 days or on nonaccrual basis) divided by total loans.	0.0480	0.0253	0.0315	0.0174
UNITB	Dummy, equals one for unit banks in states.	0.249	0.433	0.0146	0.120
LIMITB	Dummy, equals one for limited branching states.	0.593	0.491	0.475	0.499
STATEB	Dummy, equals one for statewide branching states. Excluded from the regressions as the base case.	0.150	0.357	0.501	0.500
LIMTBHC	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 states that allow statewide holding company powers.	0.556	0.497	0.481	0.500
NOINTST	Dummy, equals one for states that do not allow interstate expansions of multibank holding companies.	0.509	0.500	0.0344	0.182
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).	0.121	0.144	0.503	0.318
HERF	Deposit Herfindahl index of local market concentration.	0.227	0.164	0.233	0.155
STINC	Real state income growth (decimal)	0.0241	0.0300	0.0286	0.0231
INMSA	Dummy, equals one if the bank is in a Metropolitan Statistical Area.	0.443	0.497	0.436	0.496
FED	Dummy, equals one if the bank's primary federal regulator is the Federal Reserve.	0.0772	0.267	0.0893	0.285
FDIC	Dummy, equals one if the bank's primary federal regulator is the FDIC.	0.587	0.492	0.603	0.489
OCC	Dummy, equals one if the bank's primary federal regulator is the OCC. Excluded from the regressions as the base case.	0.336	0.472	0.308	0.461

TABLE 1, CON'T.

Symbol	Definition	1984-89 (80,014 obs.) Mean	Std. Dev.	1990-95 (65,560 obs.) Mean	Std. Dev.
SDROA	Standard deviation over past five, four, or three years (longest for which data are available) of the bank's annual return on assets.	0.00519	0.00687	0.00499	0.0117
NOSDROA	Dummy, equals one if fewer than three years of data are available with which to construct SDROA.	0.0425	0.202	0.0199	0.140

Notes: All stock values are real quantities as of the December Call Report and all prices market-averages of flows over the year divided by these stocks. The market-average price is the weighted average of the prices of the other banks in the market excluding the bank's own price, where the weights are each bank's share of the total input usage or total output production in the market. The bank's price is then the weighted average of the prices in each of the bank's markets, where the weight is the bank's share of its deposits in that market. We substitute the state-average price if the bank's deposit share in the market is over 90%. For the price of purchased funds or core deposits, we also use the state-average price if the calculated market price of purchase funds or core deposits is more than 10 percentage points over the one-year Treasury bill rate or is less than the one-year Treasury bill rate. For output prices, we use the state-average if the calculated market price is more than 25 percentage points above or is less than the one-year Treasury bill rate. We also eliminated observations in which equity was below 1% of gross total assets because the data for such banks are suspicious. 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, and MNPL. For π, an additional adjustment was made because profits can take on negative values (see text).

TABLE 2
Measured Bank X-Efficiency
U.S. Banks
(mean efficiencies, standard deviations in parentheses)

		Cost Efficiency	Standard Profit Efficiency	Alternative Profit Efficiency
1984-89	All Banks	0.806	0.567	0.558
		(0.0661)	(0.233)	(0.233)
	Number of Observations	11,040	10,997	10,982
	Small Banks (based on all-sizes frontier)	0.806	0.567	0.557
		(0.660)	(0.232)	(0.232)
	Number of Observations	10,693	10,659	10,640
	Large Banks (based on all-sizes frontier)	0.818	0.577	0.592
		(0.0684)	(0.242)	(0.251)
	Number of Observations	347	338	342
	Small Banks (based on small-bank frontier)	0.805	0.572	0.565
		(0.0656)	(0.226)	(0.225)
	Number of Observations	10,693	10,657	10,617
	Large Banks (based on large-bank frontier)	0.785	0.534	0.600
		(0.0711)	(0.222)	(0.247)
	Number of Observations	347	341	343
1990-95	All Banks	0.770	0.567	0.579
		(0.0848)	(0.229)	(0.230)
	Number of Observations	9281	9275	9267
	Small Banks (based on all-sizes frontier)	0.769	0.572	0.585
		(0.0842)	(0.223)	(0.223)
	Number of Observations	8970	8969	8960
	Large Banks (based on all-sizes frontier)	0.794	0.415	0.408
		(0.0996)	(0.337)	(0.345)
	Number of Observations	311	306	307
	Small Banks (based on small-bank frontier)	0.787	0.575	0.591
		(0.0859)	(0.221)	(0.217)
	Number of Observations	8970	8969	8959
	Large Banks (based on large-bank frontier)	0.792	0.384	0.480
		(0.0903)	(0.370)	(0.316)
	Number of Observations	311	307	302

Small banks are banks whose total assets average less than $1 billion (in 1994 dollars) over the six-year period in question and large banks are all others. In computing the means, we exclude banks whose potential profits (i.e., predicted profits if the bank were on the frontier) are less than 0.001 times mean equity for the period.

TABLE 3

Measured Bank Changes in Cost, Standard Profit, and Alternative Profit: Total Change, Productivity Change, Frontier Change, Business Conditions Change

	Cost				Standard Profit				Alternative Profit			
	Total Change	Prod. Change	Fron. Change	Bus. Cond. Change	Total Change	Prod. Change	Fron. Change	Bus. Cond. Change	Total Change	Prod. Change	Fron. Change	Bus. Cond. Change
1984-95†	0.997	1.092	1.068	0.914	1.112	1.129	1.115	0.984	1.065	1.104	1.105	0.965
1984-89†	1.014	1.061	1.037	0.955	1.240	1.278	1.303	0.970	1.097	1.142	1.133	0.960
1989-95†	0.984	1.095	1.074	0.898	1.015	1.051	1.014	0.966	1.041	1.107	1.136	0.940
1984-85	0.973	1.071	0.921	0.909	1.142	1.278	1.041	0.893	1.116	1.169	0.910	0.955
1985-86	1.001	1.068	1.095	0.937	1.662	1.847	1.971	0.900	1.894	2.072	2.619	0.914
1986-87	0.910	1.067	1.171	0.853	0.755	0.636	0.560	1.187	0.426	0.433	0.429	0.982
1987-88	1.089	1.070	0.984	1.018	1.393	1.328	1.899	1.049	1.572	1.509	1.586	1.042
1988-89	1.109	1.002	0.995	1.107	1.471	1.378	1.549	1.067	1.122	0.998	1.002	1.124
1989-90	1.043	1.049	1.169	0.994	0.088	0.098	−4.270	0.901	0.385	0.407	1.321	0.946
1990-91	0.922	1.040	1.030	0.886	9.182	9.622	‡	0.954	1.420	1.689	0.990	0.841
1991-92	0.717	1.446	1.039	0.496	1.103	0.686	0.691	1.609	1.467	1.619	1.460	0.906
1992-93	1.178	1.433	1.130	0.822	0.733	0.950	0.846	0.771	1.157	1.909	2.253	0.606
1993-94	0.999	0.987	1.009	1.012	1.902	1.403	0.972	1.355	1.815	1.373	0.782	1.322
1994-95	1.118	0.968	0.689	1.155	0.882	0.824	0.697	1.070	0.753	0.622	0.561	1.211

Total Change = for cost; $\hat{\pi}_{t+1}(X_{t+1})/\hat{\pi}_t(X_t)$ for standard profit; and $a\hat{\pi}_{t+1}(X_{t+1})/a\hat{\pi}_t(X_t)$ for alternative profit.

Prod. Change = $\hat{C}_{t+1}(X_t)/\hat{C}_t(X_t)$ for cost; $\hat{\pi}_{t+1}(X_t)/\hat{\pi}_t(X_t)$ for standard profit; and $a\hat{\pi}_{t+1}(X_t)/a\hat{\pi}_t(X_t)$ for alternative profit. For cost, a positive number indicates a decrease in productivity.

Fron. Change = $\exp[\hat{f}_{Ct+1}(X_t)]/\exp[\hat{f}_{Ct}(X_t)]$ for cost; $\{\exp[\hat{f}_{\pi t+1}(X_t)]-\theta\}/\{\exp[\hat{f}_{\pi t}(X_t)]-\theta\}$ for standard profit; and $\{\exp[\hat{f}_{a\pi t+1}(X_t)]-\theta\}/\{\exp[\hat{f}_{a\pi t}(X_t)]-\theta\}$ for alternative profit. For cost, a positive number indicates an adverse shift in the cost frontier toward higher costs.

Bus. Cond. Change = $\hat{C}_{t+1}(X_{t+1})/\hat{C}_{t+1}(X_t)$ for cost; $\hat{\pi}_{t+1}(X_{t+1})/\hat{\pi}_{t+1}(X_t)$ for standard profit; and $a\hat{\pi}_{t+1}(X_{t+1})/a\hat{\pi}_{t+1}(X_t)$ for alternative profit.

X_t = average value of X_t over all banks in the sample in year t.

†Annualized (i.e., cost or profit ratio raised to 1/(no. of years in the subperiod)).

‡Predicted standard profit in 1990 was negative.

TABLE 4

Measured Bank Changes in Cost, Standard Profit, and Alternative Profit:
Total Change, Productivity Change, Frontier Change, Business Conditions Change

Small Banks

	Cost				Standard Profit				Alternative Profit			
	Total Change	Prod. Change	Fron. Change	Bus. Cond. Change	Total Change	Prod. Change	Fron. Change	Bus. Cond. Change	Total Change	Prod. Change	Fron. Change	Bus. Cond. Change
1984-95†	0.967	1.058	1.056	0.914	1.062	1.088	1.056	0.977	1.034	1.075	1.076	0.961
1984-89†	0.989	1.036	1.040	0.954	1.073	1.106	1.082	0.970	1.050	1.096	1.076	0.958
1989-95†	0.949	1.053	1.047	0.901	1.053	1.096	1.072	0.961	1.020	1.089	1.101	0.937

Large Banks

	Cost				Standard Profit				Alternative Profit			
	Total Change	Prod. Change	Fron. Change	Bus. Cond. Change	Total Change	Prod. Change	Fron. Change	Bus. Cond. Change	Total Change	Prod. Change	Fron. Change	Bus. Cond. Change
1984-95†	0.955	1.118	1.071	0.855	1.060	1.109	1.123	0.956	1.062	1.082	1.050	0.982
1984-89†	0.992	1.058	1.015	0.937	1.061	1.056	0.764	1.005	1.089	1.062	1.106	1.026
1989-95†	0.926	1.149	1.189	0.806	1.058	1.115	1.269	0.949	1.041	1.091	1.090	0.954

In any year in question, small banks are those with total assets less than $1 billion (in 1994 dollars) and large banks are all others.

Total Change = for cost; $\hat{\pi}_{t+1}(X_{t+1})/\hat{\pi}_t(X_t)$ for standard profit; and $a\hat{\pi}_{t+1}(X_{t+1})/a\hat{\pi}_t(X_t)$ for alternative profit.

Prod. Change = $\hat{C}_{t+1}(X_t)/\hat{C}_t(X_t)$ for cost; $\hat{\pi}_{t+1}(X_t)/\hat{\pi}_t(X_t)$ for standard profit; and $a\hat{\pi}_{t+1}(X_t)/a\hat{\pi}_t(X_t)$ for alternative profit. For cost, a positive number indicates a decrease in productivity.

Fron. Change = $\exp[\hat{f}_{Ct+1}(X_t)]/\exp[\hat{f}_{Ct}(X_t)]$ for cost; $\{\exp[\hat{f}_{\pi t+1}(X_t)]-\theta\}/\{\exp[\hat{f}_{\pi t}(X_t)]-\theta\}$ for standard profit; and $\{\exp[\hat{f}_{a\pi t+1}(X_t)]-\theta\}/\{\exp[\hat{f}_{a\pi t}(X_t)]-\theta\}$ for alternative profit. For cost, a positive number indicates an adverse shift in the cost frontier toward higher costs.

Bus. Cond. Change = $\hat{C}_{t+1}(X_{t+1})/\hat{C}_{t+1}(X_t)$ for cost; $\hat{\pi}_{t+1}(X_{t+1})/\hat{\pi}_{t+1}(X_t)$ for standard profit; and $a\hat{\pi}_{t+1}(X_{t+1})/a\hat{\pi}_{t+1}(X_t)$ for alternative profit.

X_t = average value of X_t over all banks in the sample in year t.

†Annualized (i.e., cost or profit ratio raised to 1/(no. of years in the subperiod)).