WORKING PAPERS

RESEARCH DEPARTMENT

WORKING PAPER NO. 01-6/R

EXPLAINING THE DRAMATIC CHANGES IN PERFORMANCE OF U.S. BANKS: TECHNOLOGICAL CHANGE, DEREGULATION, AND DYNAMIC CHANGES IN COMPETITION

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

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

August 2002

FEDERAL RESERVE BANK OF PHILADELPHIA
Ten Independence Mall, Philadelphia, PA 19106-1574• (215) 574-6428• www.phil.frb.org

WORKING PAPER NO. 01-6/R

Explaining the Dramatic Changes in Performance of U.S. Banks: Technological Change, Deregulation, and Dynamic Changes in Competition[*,†,‡]

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

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

August 2002

Published in the *Journal of Financial Intermediation, 12 (2003), pp. 57-95*

[*]The views expressed in this paper do not necessarily represent those of the Federal Reserve Bank of Philadelphia, the Board of Governors of the Federal Reserve System, or the Federal Reserve System.

[†]We thank the editor Anjan Thakor and the anonymous referees for very helpful suggestions for improving the paper. We also thank Dennis Fixler, Diana Hancock, Dave Humphrey, Rick Lang, Mike Mohr, Leonard Nakamura, Jack Triplett, Greg Udell, Bob Yuskavage, Kim Zieschang; and participants at the Brookings Workshop on Measuring Banking Output, Washington, DC; the Conference on Service Sector Productivity and the Productivity Paradox, Ottawa, Canada; the Financial Management Association meetings; the Australian Industry Economic Conference, Canberra; and the Georgia Productivity Workshop, Athens, GA, for helpful comments and advice; Seth Bonime, Chris Malloy, Nate Miller, and Avi Peled 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) 574-4303; email: Loretta.Mester@PHIL.frb.org.

Explaining the Dramatic Changes in Performance of U.S. Banks: Technological Change, Deregulation, and Dynamic Changes in Competition*,†,‡

Abstract

We investigate the effects of technological change, deregulation, and dynamic changes in competition on the performance of U.S. banks. Our most striking result is that during 1991-1997, cost productivity worsened while profit productivity improved substantially, particularly for banks engaging in mergers. The data are consistent with the hypothesis that banks tried to maximize profits by raising revenues as well as reducing costs. Banks appeared to provide additional or higher quality services that raised costs but also raised revenues by more than the cost increases. The results suggest that methods that exclude revenues when assessing performance may be misleading.

JEL Classification Numbers: G21, G28, E58, E61, F33
Keywords: Bank, productivity, efficiency, cost, profit

Explaining the Dramatic Changes in Performance of U.S. Banks: Technological Change, Deregulation, and Dynamic Changes in Competition

1. Introduction

Much of the attention regarding the recent boom in the U.S. economy has centered on improvements in productivity associated with technological progress in information processing, telecommunications, and other technologies. Performance gains in the financial industry have also been linked to improvements in financial technologies, including new tools of financial engineering, and the more advanced use of statistical techniques. This raises the obvious question: how have these developments affected the performance of banks?

To address this broad question, we begin by noting that the banking industry may have benefitted considerably from advances in both nonfinancial and financial technologies. On the one hand, due to consolidation and deregulation, banking has become more competitive, and one would expect that many inefficiencies have been rooted out. Banks have used information processing to process deposit and loan customer information and to evaluate risks more efficiently, and telecommunications technologies to transmit this information and to process payments more quickly with fewer resources. This might suggest an improvement in cost productivity during the 1990s. On the other hand, banks have also adopted new financial technologies and new services, and improved the quality of some of the existing services. These additional services or higher service quality may have raised costs. So it may not be obvious how measured cost productivity has been affected. Similarly, it is difficult to judge *a priori* how profit productivity—which incorporates revenues as well as costs—would be affected, since this depends in part on how much more bank customers have been willing to pay for the innovations introduced by banks.

Our purpose in this paper is to examine empirically the two questions suggested by the above discussion. First, during the 1990s, has the cost productivity of the U.S. banking industry improved or worsened? Second, during the same time period, has the profit productivity of the U.S. banking industry improved or worsened?

A striking finding of this study is that during the recent 1991-1997 period, cost productivity significantly *worsened*. The predicted cost of producing a given level of output actually increased, controlling for business conditions in the local market (which include market interest rates). Another striking result in our study is that profit productivity *improved* dramatically over the same 1991-1997 period. Our findings are consistent with the hypothesis that banks provided additional services or higher service quality, which may have raised costs but also raised revenues by more than the cost increases.

Why are these additional profits not competed away? One potential explanation is that there was an increase in conventional market power in setting output prices, perhaps associated with greater market concentration after consolidation of the industry. However, some evidence presented below suggests that this is not likely the case—local market concentration changed little on average over time, and the changes that did occur are associated with price changes that contribute very little to profitability increases. A more likely explanation is that the higher profitability indicates an ongoing process of innovation. Early adopters of an effective new technology *temporarily* earn higher profits, but the abnormal returns are competed away as the technology is diffused throughout the industry. When another promising new technology comes along, the process starts again. An industry like banking that has adopted a number of successful new technologies at different points in time can have supernormal profits on average for a considerable period, although different firms may be innovating and earning the extra profits at different points in time.

Our analysis of changes in banking industry performance uses some of the recently developed concepts and techniques from the cross-section efficiency literature. The three main sources of the changes in costs and profits over time investigated are: (1) changes in managerial best practice in the industry, which includes technological change and the extent to which the best-practice firms adopt it, (2) changes in cross-section inefficiency or dispersion from this best-practice technology, and (3) changes in "business conditions," or economic factors exogenous to the banks. These first two components—changes in best practice and changes in inefficiency—together form the more traditional notion of change in *productivity*.

We analyze the sources of change in performance over the period 1984-1997 using three different optimization concepts—cost minimization, standard profit maximization, and alternative profit maximization. These concepts are based on economic optimization in reaction to market prices and competition, rather than solely on the use of technology, as are some government and research measures of productivity change. To our knowledge, only one research study has measured alternative profit performance/productivity change, no prior study has measured standard profit change, and no prior study of changes in performance or productivity has applied more than one of these concepts. We use all three concepts to ensure a comprehensive look at the data.

We decompose productivity change into the changes in the industry's best practice versus changes in inefficiency. The estimated best practice reflects the behavior of the best existing banks, but not any "true" efficient point, which would reflect the available technology and optimal responses to market prices and other business conditions. Changes in both managerial best practice and industry efficiency may be driven by technological progress, regulatory innovation, or other changes in competitive conditions. Thus, even if we rule out technological regress, productivity may worsen over time because of changes in regulation or competitive conditions. Past studies often found negative productivity growth for U.S. banks, and as indicated above, we find negative cost productivity change during the 1990s.

Section 2 gives background information on performance trends in U.S. banking. Section 3 reviews prior analysis of bank productivity change from both government statistics and research studies. Section 4 lays out the optimization concepts and how they are applied to decompose changes in performance. Section 5 gives the design of our empirical analysis. Section 6 displays our main empirical results, and Section 7 examines a number of alternative potential explanations of the empirical results. Section 8 draws conclusions.

2. Performance trends in U.S. banking

The mid-1980s to the early 1990s was a period of relatively poor performance of U.S. banks. Banks began to realize problems with commercial real estate loans and loans to less developed nations, leading to performance problems and a "credit crunch" in the early 1990s. From that time until the end of our sample period in 1997, the U.S. banking industry enjoyed substantially improved performance.

Some of these performance trends are illustrated in Table I. Profitability as measured by mean return on equity rose by more than three-quarters from 6.49% at the beginning of our sample in 1984 to 11.49% by 1997. An alternative measure of profitability, mean return on gross total assets or return on GTA, rose by almost 2/3 from 68.8 basis points to 113.4 basis points.[1] The revenues/costs ratio also rose sharply, by over 3/4, from 16.3% in 1984 to 29.4% in 1997. Nonperforming loans/total loans showed substantial improvement in the proportion of loans that were nonperforming (past due at least 90 days or on nonaccrual basis), which fell by about half, from 5.18% of loans in 1984 to 2.57% in 1997, consistent with macroeconomic improvements.

[Table I goes here]

Table I also shows that cost ratios declined dramatically over time. Mean costs/equity and costs/GTA fell by almost one-half and one-third, respectively, over the sample period. However, our estimates, discussed below, show that cost *productivity*, which takes into account the businesses conditions faced by the banks, actually worsened over this period. One of the reasons for this seeming contradiction is that the rates faced by banks to raise funds fell dramatically over time—e.g., the average interest rate faced on core deposits dropped by about 2/3 (from 6.82%in 1984 to 2.31% in 1997). Thus, bank cost ratios fell, but not as much as would be predicted by the changes in business conditions faced, particularly the market interest rates faced.

Table I also shows that most of the changes in cost, profit, and loan performance ratios occurred after 1991.

Based on these data and other factors, our main analysis focuses on explaining the changes in costs and profits over the subintervals 1984-1991 and 1991-1997, as well as over the entire 1984-1997 interval. For robustness, we also segment the data into other subintervals, examine small and large banks separately, analyze whether mergers explain the results, and assess the performance effects of industry entry and exit.

The last three columns in Table I investigate two trends that could potentially help explain part of the improvement in profit performance over time. The standard deviations of return on equity (ROE) and of return on assets (ROA) over the past three, four, or five years (the longest for which annual return data on the bank is available) are proxies for bank risk. The local deposit market Herfindahl index of concentration (HERF) is a measure of the potential for the exploitation of conventional market power. We investigate below two alternative explanations of the improvement in bank performance—an increase in risk-taking and an increase in conventional market power. The raw data shown here do not provide much support for these explanations—the standard deviations of the earnings ratios appear to have decreased over time, and the increase in local market concentration over time is quite small.

Finally, the first column of Table I shows that the industry has been consolidating rapidly, with the number of banks declining by more than 1/3 in 13 years, mostly through merger activity. Results below suggest that mergers may have played an important role in the dramatic changes in bank performance—merging banks appear to have increased costs per unit of output but made up for this by raising revenues even more.

3. <u>Prior analysis of bank performance change</u>

3.1 <u>Government productivity measures</u>

Government agencies typically measure productivity by the ratio of an output index to an input index. The U.S. Bureau of Labor Statistics (BLS) developed a labor productivity measure for the commercial banking industry (SIC 602). They measure physical banking output using a "number-of-transactions" approach based on demand deposits (number of checks written and cleared, and number of electronic funds transfers), time deposits (weighted index of number of deposits and withdrawals on regular savings accounts, club accounts, CDs, money market accounts, and IRAs), ATM transactions, loans (indexes of new and existing real estate, consumer installment, and commercial loans, and number of bank credit card transactions), and number of trust accounts, each weighted by the proportion of employee hours used in the activity. Employee labor hours are used as the denominator of the productivity index, although the BLS also computes an output per employee measure. The BLS index for banking productivity per employee hour grew at an annualized rate of 3.09% over 1984-1997, which reflects rates of change of 2.99% and 3.20% over 1984-1991 and 1991-1997, respectively. These data indicate banking productivity rising at a slower pace than the rest of the corporate sector.[2]

The Bureau of Economic Analysis (BEA) also uses labor productivity to update its output measure for banking. The BEA benchmarks the gross product originating (GPO) in banking every five years. For nonbenchmark years, real output in the banking sector is estimated by extrapolating the nominal measure of output in a benchmarking year using the rate of growth of the number of full-time equivalent employees, in effect assuming that labor productivity remains constant for five years (see Yuskavage, 1996). However, a method that estimates a decomposition of the aggregate figures has been devised, and it shows a 0.8% annual change in real output per hour over the 1989-1997 period for the Finance, Insurance, and Real Estate sector, of which banking is a part (Corrado and Slifman, 1999).

3.2 Research studies of productivity change

Several academic studies have measured productivity change. We 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.[3]

Berger and Humphrey (1992) used the thick frontier approach to compare bank cost efficiency and to study shifts in best-practice costs between 1980, 1984, and 1988 using data for virtually all U.S. banks. They found that when the shifts were not adjusted for changes in business conditions, average costs increased for all but the very largest efficient banks in the 1980-1984 interval, followed by decreases in average costs for all sizes in the 1984-1988 period. The increase in costs in the earlier period may in part reflect the deregulation of deposit rates. To the extent that the industry performed more poorly because of an increase in competitiveness that raised deposit rates, this may be a social good, because the benefits to depositors from higher rates may outweigh the higher costs to banks. When the shifts in the average cost frontiers were adjusted for changes in the business conditions, an increase in costs was still found for the 1980-1984 period, but a decrease was no longer found for the 1984-1988 period.

Bauer, Berger, and Humphrey (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 1977-1988 to estimate total factor cost productivity growth for the best-practice banks. They found an average annual growth rate of −2.28% to 0.16%, depending on the estimation method used. The poor productivity growth was attributed to higher costs of funding because of high market rates, elimination of deposit rate ceilings, and increased competition from nonbank financial intermediaries, which increased demand for funds, reduced the supply of deposits, and increased the convenience banks provided through more branches. The increase in deposit rates, increase in nonbank competition, and better convenience all made consumers 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. Measures were derived 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 (assets of $100 million-$200 million) experiencing larger increases on average than large banks. Again, much of the decline in cost productivity was attributed to deregulation of deposit rates, which has an offsetting benefit to depositors.

Again using these data, Humphrey and Pulley (1997) estimated changes in predicted profits using the alternative profit function over the 1977-1988 period and decomposed the changes that occurred after deregulation (1984-1988) into internal bank-initiated adjustments to the new regulatory structure and external changes in banks' business conditions. They found that for banks with assets over $500 million, the rise in profits from the 1977-1981 period to the 1981-1984 period resulted from a shift in the profit function and changes in business conditions, particularly deposit deregulation. Only business conditions accounted for the rise in large banks' profits from 1981-1984 to 1985-1988. For smaller banks (assets of $100 million-$500 million), there was little increase in profits between 1977-1981 and 1981-1984, and in the later period, their experience was similar to that of larger banks. The same patterns held after controlling for efficiency.

Stiroh (2000) used a panel data set of 661 top-tier bank holding companies continuously in existence during 1991-1997. He used several different specifications of outputs and several different methods of measuring cost productivity change and found small cost productivity improvements of between 0.05% and 0.47% annually. One of his specifications was similar to ours in terms of the output and input definitions, but he included many fewer variables measuring business conditions in his cost function estimations than in our specification. This is a significant difference and is likely to explain much of the difference between his results and ours. While we both find that total costs rose over the 1991-1997 period, our decomposition of this cost change between that attributable to productivity change and that attributable to a change in business conditions differs. Stiroh found that cost productivity increased thereby reducing costs, but changes in business conditions contributed to higher costs. We find that cost productivity decreased thereby raising costs, but changes in business conditions put downward pressures on costs. The difference is likely because Stiroh's business conditions include only variable outputs, fixed netputs, and input prices. In contrast, in addition to these, we control for a number of additional business conditions in the market, including state income growth, market nonperforming loans, the extent of interstate branching, urban vs. rural market, market concentration, and federal regulator. Some of these conditions, like nonperforming loans and state income growth, were generally improving over the period and likely improved bank performance in a number of ways, including fewer costs expended dealing with problem loans. To the extent that these improved conditions are exogenous to the bank, we would not want to conclude that improved performance derived from these conditions is a productivity improvement for the bank. Rather, we would want to label this as an improvement in the exogenous business conditions faced by the bank. In contrast, Stiroh's methodology would label this as an improvement in bank productivity.

Several research efforts used linear programming methods to measure changes in productivity. These methods are nonstochastic and do not allow for random error. The productivity changes are based on quantities of outputs and inputs without regard to prices, so 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.

Devaney and Weber (2000) investigated whether the market structure of rural banking markets affected productivity growth over 1990-1993. They used linear programming 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. They found positive productivity growth at rural banks over 1990-1993. Shifts in the production frontier were the driving force of this productivity growth.

Wheelock and Wilson (1999) also used linear programming and decomposed the change in productivity into its change in efficiency and frontier shift components. While banks on the frontier improved over the period 1984-1993, productivity declined on average during this period because of reductions in efficiency. Most banks, particularly smaller banks (assets below $300 million), were not able to adapt quickly to changes in technology, regulations, and competitive conditions and fell further away from the efficient frontier.

Similarly, Alam (2001) applied linear programming to a balanced panel of 166 banks with greater than $500 million in assets and uninterrupted data from 1980 to 1989. She found that productivity surged between 1983 and 1984, retreated over the next year, and grew again between 1985 and 1989. The main source of the productivity growth was a shift in the frontier rather than a change in efficiency.

4. The optimization concepts and the decomposition of cost and profit changes

4.1 Cost minimization

The cost minimization concept assumes that firms minimize variable costs subject to exogenously given prices of variable inputs, quantities of variable outputs, quantities of fixed netputs (fixed inputs or outputs), environmental factors, their own managerial inefficiency, and random error. This concept is implemented using a standard cost function that relates variable costs to these exogenously given conditions. For simplicity, the inefficiency and random error are assumed to be multiplicatively separable from the rest of the cost function, and all of the variables (other than dummies) are measured in natural logs:

$$lnC \;=\; f_C(X_C) + ln u_C \; + ln\epsilon_C. \tag{1}$$

The variable lnC measures log of variable costs (including both operating and interest expenses); $f_C(\cdot)$ is the best-practice (log) cost function; $X_C \equiv (lnw, lny, lnz, lnv)$ is the set of logged exogenous "business conditions" that affect costs, specifically, variable input prices (lnw), variable output quantities (lny), fixed netput quantities (lnz), and environmental variables (lnv).[4] The lnu_C term denotes an inefficiency factor that is zero for best-practice firms and raises costs for other firms. The $ln\epsilon_C$ term is a random error assumed to have zero mean each period.

We represent the cost of the industry at time t by the predicted cost of a bank with average business conditions, average inefficiency for the period, and a zero random error. This gives $\exp[f_{Ct}(\bar{X}_{Ct})] \bullet \exp[ln\bar{u}_{Ct}]$, where $\bar{X}_{Ct}$ gives the average values of the business condition regressors at time t and $ln\bar{u}_{Ct}$ gives the average value of the inefficiency factor. The total gross change in cost between period t and period t+k is measured by the ratio of the predicted costs in the two periods:

$$\Delta TOTAL_{Ct,t+k} \;\equiv\; \{\exp[f_{Ct+k}(\bar{X}_{Ct+k})] \bullet \exp[ln\bar{u}_{Ct+k}]\} \,/\, \{\exp[f_{Ct}(\bar{X}_{Ct})] \bullet \exp[ln\bar{u}_{Ct}]\}. \tag{2}$$

As this is a gross change, a number below 1 indicates falling costs, and a number above 1 indicates rising costs. All data are measured in 1994 dollars, so we are measuring real changes in costs. For example, a finding of 1.05 indicates that real costs have increased by 5% between t and t+k. To make the findings easier to follow, the tables will report the annualized average rate of change over the interval, i.e., the kth root of the k-period rate of change [e.g., $(\Delta TOTAL_{Ct,t+k})^{1/k}$].

We decompose $\Delta TOTAL_C$ into the gross changes in best practice, inefficiency, and business conditions:

$$\Delta TOTAL_{Ct,t+k} \;=\; \{\exp[f_{Ct+k}(\bar{X}_{Ct})] \,/\, \exp[f_{Ct}(\bar{X}_{Ct})]\} \;\bullet \qquad \text{(Change in best practice)}$$

$$\{\exp[ln\bar{u}_{Ct+k}] / \exp[ln\bar{u}_{Ct}]\} \bullet \qquad \text{(Change in inefficiency)}$$

$$\{\exp [f_{Ct+k}(\bar{X}_{Ct+k})] / \exp [f_{Ct+k}(\bar{X}_{Ct})]\} \qquad \text{(Change in business conditions)}$$

$$\equiv \Delta BESTPR_{Ct,t+k} \bullet \Delta INEFF_{Ct,t+k} \bullet \Delta BUSCOND_{Ct,t+k} . \tag{3}$$

Thus, the change in costs is decomposed into three multiplicative terms. The change in best practice, $\Delta BESTPR_C$, gives the change in costs due to changes in the best practice cost function $f_C(\cdot)$, since it holds business conditions and inefficiency constant. Similarly, $\Delta INEFF_C$ and $\Delta BUSCOND_C$ give the contributions from changes in inefficiency and business conditions only, respectively. All three terms are measured as gross changes.

Cost productivity change is the product of the change in best practice and the change in inefficiency:

$$\Delta PROD_{Ct,t+k} \equiv \Delta BESTPR_{Ct,t+k} \bullet \Delta INEFF_{Ct,t+k}$$

$$= \{\exp [f_{Ct+k}(\bar{X}_{Ct})] / \exp [f_{Ct}(\bar{X}_{Ct})]\} \bullet \{\exp[ln\bar{u}_{Ct+k}] / \exp[ln\bar{u}_{Ct}]\}. \tag{4}$$

Although change in best practice and the change in inefficiency are different concepts, it is informative to combine them into a single measure of productivity change, since this concept has been used in prior research and is reported in government statistics. $\Delta PROD_C$ is also relatively easy to estimate, whereas dividing it into $\Delta BESTPR_C$ and $\Delta INEFF_C$ is likely to involve more estimation error.

Cost productivity change, $\Delta PROD_C$, represents an improvement over the government statistics. As discussed, the government measures use the change in a single output, such as a weighted sum of bank transactions, divided by the quantity of a single input measure, employee labor hours. $\Delta PROD_C$ is a superior indicator of productivity in our opinion because it controls for all of the output quantities as well as the input prices, fixed netput quantities, and environmental conditions specified in the business conditions vector $\bar{X}_C$. It is important to control for these factors, so that a change in costs that is not due to any decision or managerial skill of the bank is not attributed to a change in productivity.

$\Delta PROD_C$ also includes all variable costs, including non-labor physical input costs, other noninterest expenses, and interest costs, rather than just employee labor hours, as in the government statistics. Bank employee labor hours may also be an inaccurate indicator of labor input because of a trend toward outsourcing some operations to holding company affiliates and service bureaus, so that the change in output per employee hour may overstate the change in output per total labor hour worked by employees and nonemployees. As of 1997 (a benchmark year for the BEA), labor compensation expenses accounted for only 32.0% of the BEA's gross domestic product of depository institutions (BEA, December 2000). This figure explicitly excludes the compensation and product of employees working elsewhere in bank holding companies. It has been shown that the ratio of the number of employees to costs has declined dramatically over time (Berger and Humphrey, 1992), as bank holding companies have moved many of their back-office operations outside the bank itself. Thus, costs are incurred at the bank level and are measured in other noninterest expenses component of costs in the Call Report, but this labor is not measured for the bank. Failure to account either for the labor used elsewhere in the holding company but effectively working for the bank or for the cost of this labor and capital could bias government productivity measures toward a spurious finding of productivity improvement. Interest expenses on purchased funds also often represent physical inputs involved in raising the funds at the institutions from which the funds were purchased, and so should be included in our opinion. In addition, it is important to include interest expenses on deposits because banks often substitute between spending additional real resources to provide service and paying higher rates on deposits.

4.2 Standard profit maximization

The two profit maximization concepts assume that firms maximize variable profits, again subject to exogenous business conditions. Standard and alternative profit maximization differ from one another only in terms of the specification of business conditions. In studying firm performance, profit maximization is superior to cost minimization because it more completely describes the economic goals of managers and owners, who take revenues into account as well as costs. For example, a decision that raises both revenues and costs, but raises revenues by more than it raises costs, will appropriately be counted as an improvement in performance under profit maximization, but may be counted as a deterioration under cost minimization.[5] The nonparametric methods and government productivity statistics also generally neglect the beneficial effects of revenue gains.

Standard profit maximization is implemented using a profit function that specifies output prices in the business conditions vector in place of the output quantities specified in the cost function, but all other business conditions remain the same. Thus, firms are assumed to choose their outputs in response to relative output prices and other factors in the maximization process. The standard profit function is given by:

$$ln(\pi + \theta) = f_{\pi}(X_{\pi}) + ln u_{\pi} + ln\epsilon_{\pi}, \tag{5}$$

where π is the variable profits of the firm, which includes all interest and fee income earned on variable outputs minus variable costs C, which is used in the cost function. Because profits may be negative, the same scalar θ is added to every firm's dependent variable in a given time period before logging, so that the log is taken of a positive number (θ varies over time). $X_\pi \equiv (ln\text{w}, ln\text{p}, ln\text{z}, ln\text{v})$ is the same as X_C , except logged output prices *ln*p replace logged output quantities *ln*y. Analogous to the case of the cost function, $f_\pi(\cdot)$ is the best-practice profit function, $ln\text{u}_\pi$ is an inefficiency factor that is zero for best-practice firms and negative for other firms, reducing their profits below the best-practice level, and $ln\epsilon_\pi$ is a random error with a mean of zero each period.

The decomposition of the change in profit over time is similar to the cost minimization case, but the formulas differ slightly because of the nonlinearity introduced by θ. The representative profit of a bank with average business conditions, average inefficiency for the period, and a zero random error at time t is given by $\exp[f_{\pi t}(\bar{X}_{\pi t})] \bullet \exp[\overline{ln\text{u}}_{\pi t}] - \theta_t$, and the total gross change in profit between periods t and t+k is given by:

$$\Delta \text{TOTAL}_{\pi t,t+k} = \{\exp[f_{\pi t+k}(\bar{X}_{\pi t+k})] \bullet \exp[\overline{ln\text{u}}_{\pi t+k}] - \theta_{t+k}\} / \{\exp[f_{\pi t}(\bar{X}_{\pi t})] \bullet \exp[\overline{ln\text{u}}_{\pi t}] - \theta_t\}. \tag{6}$$

Here, a figure above 1 indicates an improvement in profits, so that a figure of 1.05 would indicate that profits have increased or improved by 5% between t and t+k. The components of $\Delta\text{TOTAL}_{\pi t,t+k}$ are decomposed as:

$$\Delta \text{BESTPR}_{\pi t,t+k} = \{\exp [f_{\pi t+k}(\bar{X}_{\pi t})] - \theta_{t+k}\} / \{\exp [f_{\pi t}(\bar{X}_{\pi t})] - \theta_t\}$$

$$\Delta \text{INEFF}_{\pi t,t+k} = \langle \{\exp[f_{\pi t+k}(\bar{X}_{\pi t+k})] \bullet \exp[\overline{ln\text{u}}_{\pi t+k}] - \theta_{t+k}\} / \{\exp[f_{\pi t}(\bar{X}_{\pi t})] \bullet \exp[\overline{ln\text{u}}_{\pi t}] - \theta_t\} \rangle / \langle \{\exp[f_{\pi t+k}(\bar{X}_{\pi t+k})] - \theta_{t+k}\} / \{\exp[f_{\pi t}(\bar{X}_{\pi t})] - \theta_t\} \rangle$$

$$\Delta \text{BUSCOND}_{\pi t,t+k} = \{\exp [f_{\pi t+k}(\bar{X}_{\pi t+k})] - \theta_{t+k}\} / \{\exp [f_{\pi t+k}(\bar{X}_{\pi t})] - \theta_{t+k}\} . \tag{7}$$ [6]

4.3 Alternative profit maximization

Alternative profit maximization has the same objective as the standard profit maximization concept, but specifies the same set of business conditions as under cost minimization—the logged output quantities *ln*y are specified in the X vector, rather than logged output prices *ln*p. The alternative profit function is given by:

$$ln(\pi + \theta) = f_{a\pi}(X_C) + ln\text{u}_{a\pi} + ln\epsilon_{a\pi}. \tag{8}$$

The total gross change in alternative profit $\Delta\text{TOTAL}_{a\pi}$ will be the same as the total gross change in standard profit ΔTOTAL_π (the gross change in average variable profits), but the decompositions into the various components will differ because of the use of the slightly different business conditions.

We do not believe that firms actually take their outputs as given and maximize profits, as the alternative profit specification literally implies. We would not use the alternative profit maximization concept if the assumptions behind the cost minimization and standard profit maximization concepts held precisely. Nonetheless, Berger and

Mester (1997) identified four violations of these assumptions under which the alternative profit concept may provide useful information in efficiency measurement, and we apply them here in terms of measuring performance/productivity changes over time.

First, if there are substantial unmeasured changes in the quality of banking services over time, and customers are willing to pay more for higher quality, banks should receive higher revenues that compensate for their extra costs of producing high quality. The cost measures may treat an unmeasured improvement in quality over time as a deterioration in performance, whereas alternative profit measures take into account the extra revenues that cover these costs. Second, the variable outputs may not be completely variable, as assumed by the standard profit concept. If there are increases over time in scale economies in banking, and banks cannot adjust their size quickly, then the standard profit approach may find inefficiency increasing over time as banks fall further below efficient scale. The alternative profit approach may partially mitigate this problem by simply evaluating bank performance at their existing output levels. Third, banks may have some market power over the prices of their outputs, contrary to the standard assumption of exogenous prices. An increase in the exercise of market power that raises prices over time may be measured as an exogenous improvement in business conditions when applying the standard profit concept, but may be measured as an improvement in best practice when applying the alternative profit concept, neither of which is precisely correct. Fourth, if output prices are not accurately measured, as is generally the case in banking research (including this study), the standard profit function may be inaccurately measured, resulting in inaccurate measurement of $\Delta BESTPR_{\pi}$, $\Delta INEFF_{\pi}$, and $\Delta BUSCOND_{\pi}$. The alternative profit function may provide an alternative measurement of the components of the change in profitability that does not depend on output prices to check robustness.[7] Because one or more of the assumptions underlying the cost minimization and standard profit maximization concepts are likely to be violated by the data, and because we wish to be comprehensive, we apply all three optimization concepts.

5. Methodological design

Our data set is primarily drawn from the Reports of Income and Condition (Call Reports). For each year from 1984 through 1997, the data set includes annual information on virtually all U.S. commercial banks that operated in the year, although we primarily focus on data from 1984, 1991, and 1997. Because of industry consolidation, the number of observations declines from 14,095 in 1984 to 11,623 in 1991 to 8,855 in 1997, as shown in Table I above.

5.1. Variables

Table II gives the definitions of the variables in the cost and profit functions, their sample means, and standard deviations for 1984, 1991, and 1997. Although the continuous variables are generally expressed in natural logs in the cost and profit functions, we show means and standard deviations of the levels to be more informative.

In choosing which financial accounts to specify as outputs versus inputs, we use the "asset approach" or "intermediation approach" (Sealey and Lindley 1977). All liabilities (core deposits and purchased funds) and financial equity capital provide funds and are treated as inputs, and all assets (loans and securities) use bank funds and are treated as outputs.[8] Physical inputs (labor and premises) are specified as inputs that generate costs.

[Table II goes here]

For the input and output prices, we specify the market-average price **faced**, rather than the actual price paid or received by the bank. As described in the notes to Table II, only data from other firms in the bank's local markets are used to construct these market-average prices. The market-average prices faced are more likely to be exogenous to the bank than the prices actually paid or received by the bank. A second advantage is that any mistakes the bank makes in setting prices for its inputs or outputs given the market price conditions will be counted properly as inefficiencies, rather than just high or low prices or good or bad business conditions. For example, a bank that sets its deposit rate well above those of its market competitors, all else equal, will be measured as inefficient given its market-average deposit rate faced in the *ln*w vector. Market-average prices are also likely to average out some of the computational errors in measuring prices of individual banks.

The variable inputs for which prices *ln*w are specified are purchased funds, core deposits, and labor. The variable outputs *ln*y are 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 included. We specify off-balance-sheet items, physical capital, and financial equity capital as fixed netputs *ln*z. For the off-balance-sheet items, we use the Basel Accord risk weights on the assumption that the output may be roughly proportional to the perceived credit risk on which these weights are based. We specify these items as fixed primarily because of the difficulty of obtaining accurate price information. We also treat physical capital (premises and equipment) as a fixed input because it is slow to adjust and because it is difficult to measure prices for these durable inputs. Financial equity capital is an input under the asset approach, which we treat as fixed, in part because it is difficult to change quickly and in part because its price (the risk-adjusted expected return on equity) is difficult to measure. In addition, banks must meet regulatory capital requirements that may not be consistent with cost minimization or profit maximization. It is important to include equity because it directly affects other costs and is an alternative source of funding for bank assets. It may affect the risk premium a bank pays for purchased funds, since equity provides a cushion against insolvency and an incentive to control risks (Hughes and Mester, 1998).[9]

Among the environmental variables *ln*v we include the log of the market-average nonperforming loans to total loans ratio, *ln*MNPL, and ½(*ln*MNPL)^2, since dealing with exogenous loan problems raises costs and lowers profitability. We use the market average rather than the individual bank's ratio, since the market average captures the exogenous conditions in markets that affect loan performance. Market conditions are also accounted for by state income growth (STINC, ½STINC^2). We specify controls for the state geographic restrictions on bank competition, including unit banking (UNITB), limited branching (LIMITB), with statewide branching as the base case; the degree of in-state holding company expansion permitted (LIMTBHC); whether out-of-state holding company expansion is prohibited (NOINTST); and the proportion of the U.S. banking assets held in states allowed to enter the bank's own state (ACCESS, ½ACCESS^2). We also include the Herfindahl index of local deposit market concentration (HERF); whether the bank is located in a metropolitan area (INMSA); and the identity of a bank's primary federal regulator (FED, FDIC, with OCC as the base case).

5.2 Functional form

We use the Fourier-flexible functional form, a global approximation that includes a standard translog plus Fourier trigonometric terms. Our specification of the cost function is:

$$\begin{aligned} ln(C/w_3z_3) = {} & \alpha + \Sigma_{i=1}^{2} \beta_i\, ln(w_i/w_3) + \tfrac{1}{2} \Sigma_{i=1}^{2} \Sigma_{j=1}^{2} \beta_{ij}\, ln(w_i/w_3)\, ln(w_j/w_3) + \Sigma_{k=1}^{4} \gamma_k\, ln(y_k/z_3) \\ & + \tfrac{1}{2} \Sigma_{k=1}^{4} \Sigma_{m=1}^{4} \gamma_{km}\, ln(y_k/z_3)\, ln(y_m/z_3) + \Sigma_{r=1}^{2} \delta_r\, ln(z_r/z_3) + \tfrac{1}{2} \Sigma_{r=1}^{2} \Sigma_{s=1}^{2} \delta_{rs}\, ln(z_r/z_3)\, ln(z_s/z_3) \\ & + \Sigma_{i=1}^{2} \Sigma_{k=1}^{4} \eta_{ik}\, ln(w_i/w_3)\, ln(y_k/z_3) + \Sigma_{i=1}^{2} \Sigma_{r=1}^{2} \rho_{ir}\, ln(w_i/w_3)\, ln(z_r/z_3) + \Sigma_{k=1}^{4} \Sigma_{r=1}^{2} \tau_{kr}\, ln(y_k/z_3)\, ln(z_r/z_3) \\ & + \Sigma_{k=1}^{4} [\phi_k \cos(q_k) + \omega_k \sin(q_k)] + \Sigma_{k=1}^{4} \Sigma_{m=k}^{4} [\phi_{km} \cos(q_k+q_m) + \omega_{km} \sin(q_k+q_m)] \\ & + \Sigma_{k=1}^{4} [\phi_{kkk} \cos(q_k+q_k+q_k) + \omega_{kkk} \sin(q_k+q_k+q_k)] + \Sigma_{n=1}^{14} \xi_n\, ln v_n + ln u_C + ln \epsilon_C\,, \end{aligned} \tag{9}$$

which is estimated separately for each year, allowing all the parameters to vary. The variables (y_k/z_3), (z_r/z_3), and MNPL (one of the environmental variables *ln*v) have 1 added before logging for every firm to avoid taking the log of zero. The q_k terms are rescaled values of the $ln(y_k/z_3)$, such that each of the q_k is in the interval $[0,2\pi]$, where π here refers to the number of radians (not profits).[10] The standard symmetry restrictions apply to the translog portion of the function (i.e., $\beta_{ij} = \beta_{ji}$, $\gamma_{km} = \gamma_{mk}$, $\delta_{rs} = \delta_{sr}$). We do not include 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 inefficiency (i.e., no errors in responding to relative prices).[11]

The standard and alternative profit functions use essentially the same specification as the cost function with a few changes. First, the dependent variable for the profit functions replaces $ln(C/w_3z_3)$ 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, $\theta_t \equiv |(\pi/w_3z_3)_t^{min}| + 1$ is added to every firm's dependent variable so that the natural log is taken of a positive number, since the minimum profits are typically negative.[12] For the alternative profit function, this is the only change in specification (other than relabelling 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 translog terms containing the variable output quantities, $ln(y_k/z_3)$, are replaced by the corresponding output prices, $ln(p_k/w_3)$, and the trigonometric terms containing the output quantities q_k are dropped.

As shown, the cost, profit, and price terms are normalized by the last input price, the price of labor w_3, in order to impose linear homogeneity on the models.[13] We also normalize the cost, profit, output quantities, and fixed netput quantities by the last fixed netput, financial equity capital z_3. Since the costs and profits of the largest firms are many times larger than those of the smallest firms, large firms would have random errors with much larger variances in the absence of the normalization, and division by equity should drastically reduce this heteroskedasticity. This normalization may also help reduce a bias toward finding high standard profit efficiency for the largest banks, since these banks may 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. Division by equity may also give the dependent variables more economic meaning—the profit dependent variables become essentially the bank's return on equity, or ROE (normalized by prices and with a constant added), a commonly accepted measure of how well the bank is using its scarce financial capital.

5.3 Methods used to decompose the total changes in costs and profits

We decompose the total changes in costs and profits over time, $\Delta TOTAL_{Ct,t+k}$, $\Delta TOTAL_{\pi t,t+k}$, and $\Delta TOTAL_{a\pi t,t+k}$ in several steps. First, we estimate simple average-practice cost and profit functions for each year that include all banks whether they use best-practice versus inefficient techniques. The change in the average-practice functions over time reflects both the change in best practice (i.e., the change in $f(\cdot)$) and the change in inefficiency (i.e., the change in $\overline{ln u}$). We then use these average-practice functions to separate ΔTOTAL into the productivity change ΔPROD and the change in business conditions ΔBUSCOND components.

The changes in $\Delta PROD_{Ct,t+k}$, $\Delta PROD_{\pi t,t+k}$, and $\Delta PROD_{a\pi t,t+k}$ are the gross changes from period t to t+k in costs or profits from evaluating the average-practice function, holding business conditions constant at their period t levels. To see this, we rearrange the terms for $\Delta PROD_{Ct,t+k}$ from Eq. (4) above to give:

$$\Delta PROD_{Ct,t+k} = \{\exp[f_{Ct+k}(\bar{X}_{Ct})] / \exp[f_{Ct}(\bar{X}_{Ct})]\} \bullet \{\exp[\overline{ln u}_{Ct+k}] / \exp[\overline{ln u}_{Ct}]\}$$
$$= \{\exp[f_{Ct+k}(\bar{X}_{Ct})] \bullet \exp[\overline{ln u}_{Ct+k}]\} / \{\exp[f_{Ct}(\bar{X}_{Ct})] \bullet \exp[\overline{ln u}_{Ct}]\} . \qquad (10)$$

The numerator is the predicted cost from the average-practice cost function from period t+k applied to the business conditions data from period t, and the denominator uses period t information for both the average-practice cost function and business conditions data. Put another way, cost productivity change includes the changes in best practice and inefficiency, both of which are incorporated in the average-practice cost function and holds business conditions unchanged. We simply use the estimated parameters of average-practice cost functions to evaluate the numerator and denominator of (10). We estimate the changes in costs or profits that are due to changes in business conditions as the changes in costs or profits that remain after accounting for the productivity changes (e.g., $\Delta BUSCOND_{Ct,t+k}$ is given by $\Delta TOTAL_{Ct,t+k} / \Delta PROD_{Ct,t+k}$).

In general, the predicted costs or profits from the average-practice function evaluated at the average business conditions for that year and a zero random error term will not precisely equal the average total cost or profit for the industry because the transformations of the dependent variables to estimate the levels of costs or profits are nonlinear. We correct for this by multiplying the predicted levels from every average-practice cost or profit function by a constant, such that the predicted cost or profit at the mean value of business conditions for the year in which the function was estimated equals the sample average cost or profit for that year. In this way, our average-practice functions correctly predict the ΔTOTAL that they are being used to decompose, and the estimated ΔPROD and ΔBUSCOND correctly multiply to ΔTOTAL.

Finally, we decompose the productivity changes ΔPROD into the change in best practice ΔBESTPR and change in inefficiency ΔINEFF components. There is no consensus as to the best way to estimate the best-practice frontier. We use a version of the thick frontier method to measure ΔBESTPR (Berger and Humphrey, 1991). For each year, we divide up the banks based on their residuals from estimating the average-practice cost and profit functions. Banks with residuals in the "best" 25% in each of ten size categories (i.e., lowest cost residuals or highest profit residuals for their size category) are assumed to be best practice for that year. We then estimate the best-practice cost and profit functions using OLS on this most efficient quarter of banks. These estimated thick frontiers are treated as the best-practice functions $f_C(\cdot)$, $f_\pi(\cdot)$, and $f_{a\pi}(\cdot)$.[14] The ΔBESTPR are measured as the changes in costs or profits due to changes in f(·), holding business conditions constant at period t values. The changes in inefficiency ΔINEFF are estimated as the changes in productivity ΔPROD that remain after accounting for the change in best practice ΔBESTPR (e.g., $\Delta INEFF_{Ct,t+k}$ is estimated as $\Delta PROD_{Ct,t+k}$ / $\Delta BESTPR_{Ct,t+k}$). Because of the uncertainty involved in the estimation of the thick frontier, the breakout of the change in productivity into its components should be considered less accurate than the other decompositions.

6. Main empirical results–Cost and profit changes for all U.S. banks, 1984-1997

The top panel of Table III reports the total changes in costs and profits (ΔTOTAL), and the decompositions of these total changes into their ΔPROD, ΔBUSCOND, ΔBESTPR, ΔINEFF components for all U.S. banks over 1984-1997, and over the two subintervals 1984-1991 and 1991-1997.

[Table III goes here]

The $\Delta TOTAL_C$ figures show that the cost of the average bank rose at an annual rate of 1.1% over the entire 1984-1997 interval, falling at an annual rate of 0.3% over the first seven years from 1984 to 1991, and rising at an annual rate of 2.7% over the subsequent six years from 1991 to 1997. These trends differ from those of the cost ratios shown in Table I above, which were scaled by equity or GTA. Here, we include these scale factors in our cost business conditions vector X_C. Financial equity capital is explicitly included in X_C as the third fixed netput (z_3), and GTA is implicitly included in X_C through the inclusion of its components, the asset output quantities plus the physical capital fixed netput $((\sum_{k=1}^{4} y_k) + z_2)$.

Using the average-practice cost function (estimated using all banks) to decompose the cost changes suggests that cost productivity worsened over both subintervals ($\Delta PROD_C > 1$), while the business conditions as a whole reduced costs over both subintervals ($\Delta BUSCOND_C < 1$). Moreover, these changes are accentuated and quite substantial in the 1991-1997 subinterval, with measured changes in productivity increasing costs at an annual rate of 12.5% and measured changes in business conditions lowering costs at an annual rate of 8.7%.

The strong benefits of these changes in business conditions in lowering costs are not at all surprising. As shown in Table II, interest rates on purchased funds and core deposits declined substantially during both subintervals. Given that interest expenses make up more than half of variable costs, it is expected that these declines in rates would reduce costs substantially. Somewhat offsetting these declines was the increase in the price of labor. The market-average price of labor, w_3, rose from \$30.7 thousand in 1984 to \$32.5 thousand in 1991 to \$36.2 thousand in 1997. Financial equity capital (z_3) grew by 5.1% on an annualized basis from 1984 to 1991 and 12.3% from 1991 to 1997, and GTA ($(\sum_{k=1}^{4} y_k) + z_2$) grew by 3.7% and 8.3% per annum over the two subintervals, which helps explain why the cost ratios shown in Table I declined. As to the environmental variables, the decline in market nonperforming loans likely reduced the costs of dealing with these loans significantly, but the effects of liberalizations of state geographic restrictions on competition may either improve or worsen measured cost performance, as discussed further below.

More puzzling are the measured unfavorable shifts in cost productivity of 4.2% of costs annually over 1984-1991 and 12.5% of costs annually over 1991-1997. Using the best-practice cost function (i.e., using the "best" 25% of banks in each of ten size classes) suggests that both best-practice costs and inefficiency worsened in both subintervals and that an unfavorable shift in best practice explains most of the worsened productivity over the 1991-1997 subinterval (9.3 percentage points of the 12.5 percentage points). As noted, this breakout of the change in productivity into its components should be considered less accurate than the other decompositions, because of the uncertainty involved in the estimation of the thick frontier. Nonetheless, robustness checks using the best 15% and best 35% in each size class in place of the best 25% yielded qualitatively similar results.

These cost productivity/best practice deteriorations over time are in sharp contrast to the BLS annualized labor productivity improvements of 2.99% over 1984-1991 and 3.20% over 1991-1997, although other research studies of productivity change often found productivity deteriorations. Unlike the government statistics, our cost productivity measure controls for all of the output quantities, input prices, fixed netput quantities, and environmental conditions specified in the business conditions vector X_C. Our variable costs include noninterest expenses, which incorporate costs incurred elsewhere in the holding company in support of the bank.

Previous research cited the deregulation of deposit interest rates in the early 1980s and the resulting disequilibrium as important factors in worsening cost performance, but these factors almost surely cannot explain our worsening cost productivity/best-practice over 1991-1997, which is well after the deregulation of the rates. There may have been an increase in competition owing to liberalization of state geographic restrictions on competition or to exogenous market developments, which may make productivity/best practice appear to worsen because of a measurement problem in which the benefits to consumers are not taken into account in the cost function. However, it seems likely that these effects would be concentrated in the states that liberalized their rules the most and would largely be captured in the business conditions variables.[15]

The profit figures in the top panel of Table III show a dramatically different picture from the cost figures. The profit of the average bank rose at an annual rate of 7.9% over 1984-1997, and at annual rates of 4.3% and 12.2% over 1984-1991 and 1991-1997, respectively. The data suggest that the increase in profits is entirely due to improvements in profit productivity—measured business conditions actually lowered profits slightly. The standard and alternative profit results differ as to whether the improvement in productivity is due to both improvements in best practice and better efficiency (standard profit result) or whether it is essentially all due to improvements in best practice (alternative profit result). Most important and striking, both profit approaches find substantial increases in productivity during 1991-1997—annual increases of 13.7% and 16.5% for standard and alternative profit, respectively—in sharp contrast to the cost productivity deterioration for this subinterval.

7. Potential explanations of the empirical results

7.1 Revenue-based productivity gains

This section tries to explain the striking finding that cost productivity worsened while profit productivity improved substantially over the 1991-1997 period. One likely explanation is revenue-based productivity gains. The cost approach and other approaches that do not consider revenues may not capture unmeasured changes in output quality over time or the profit maximization goal of banks, which requires that effort be spent to raise revenues as well as reduce costs. Efforts to raise quality may be measured as deteriorations in cost performance because these efforts raise expenditures, even if they increase revenues by more than the cost increases. The nonparametric concepts and the government statistics suffer from these weaknesses as well. Use of the profit approaches may help take into account unmeasured changes in the quality of banking services by including higher revenues paid for the improved quality, and may help capture the profit maximization goal by including both the costs and revenues.

Over time, banks have offered wider varieties of financial services, such as mutual funds, derivatives, and other products of financial engineering that help customers invest and manage risks. In addition, banks have provided additional convenience through more extensive branching and ATM networks, expanded availability of debit and credit cards, and a proliferation of online services. It seems likely that providing these new services and improving service quality significantly increased costs, but banks presumably made the necessary expenditures to maximize profits through higher prices or expanding or maintaining market shares. Our cost and profit productivity results are consistent with this hypothesis.

As noted above, the supernormal profits from adopting new technologies may persist over time in the industry if there is a series of innovations, as appears to have occurred in the banking industry. The early adopters of a successful new nonfinancial and financial technology earn supernormal profits from higher prices or increased market share, and others follow to avoid losing market share. The supernormal profits are competed away as the innovations become widely adopted. However, as long as the innovation process is ongoing, and different new technologies are introduced at different points in time, innovating banks can keep earning high profits on different new or improved products. The early adopters of the series of new technologies may earn sufficient additional profits to raise the industry average for a considerable period, although the early adopters of different technologies need not consistently be the same banks.

As an example of how technology takes time to become adopted, consider small-business credit scoring, a set of statistical tools used to quantify and price risks. Prior to the mid-1990s, credit scoring was widely used in consumer lending, but it was little used in small-business lending for lack of a large loan database to develop a reliable model. Since 1995, however, small-business credit scoring models have become available from outside vendors that pool data from a number of banks (Mester, 1997).[16] Credit scoring has been shown to be successful in increasing the small business loan market shares for the banks that have adopted it. Nonetheless, this technology is still in the process of adoption. As of January 1998, 37% of a sample of the largest banks in the U.S. still had not adopted small-business credit scoring, and this technology is still not used by most small banks (Frame, Srinivasan, and Woosley, 2001).

7.2 Selection of time intervals

We next consider the possibility that our findings may be driven by the way we have divided up the 1984-1997 period into two subintervals. To check for robustness, the bottom panel of Table III repeats the cost and profit change analysis of the top panel except the data are segmented into three intervals, 1984-1989, 1989-1992, and 1992-1997. The year 1989 was a relatively poor one for U.S. bank performance, and the 1989-1992 subinterval generally corresponds to what most consider to be the "credit crunch" period.[17]

The results shown in the bottom panel of Table III suggest that most of the main results are robust. Again, cost productivity worsens in the early and late time subintervals and again profit productivity improves over time, with cost worsening more and profits improving more in the 1990s than in the 1980s. However, the credit crunch subinterval 1989-1992 does differ in that cost productivity and best practice improve, while profit productivity and best practice improve much more than over any other subinterval. These unusual findings are reversed when the 1992-1997 subinterval is included along with the 1989-1992 subinterval—the worsening of cost productivity and best practice over 1992-1997 would more than overwhelm the improvements of 1989-1992. Overall, these data remain consistent with the hypothesis that banks' goals over time were to maximize profits and that they provided additional services that raised costs but raised revenues even more. However, these data also suggest that the increased costs from producing higher-quality services may be temporarily overwhelmed by extreme improvements in cost productivity.

7.3 Heterogeneity between small and large banks

Another potential candidate for explaining our results is the inclusion of both small and large banks in the same analysis. Small and large banks may provide different products that require use of different technologies. For example, a $1 billion loan issued by a large bank may be a different product requiring different monitoring and screening techniques than 10,000 loans of $100,000 each that may be issued (in aggregate) by small banks, but the balance sheet entries do not distinguish among these products. Also, it is likely that small banks and large banks adopt new technologies at different rates, and adapt to deregulation in different ways. To some extent, the observed effects thus far could reflect the shifting of banks into larger size classes over time, which may use different techniques to produce different products having different cost and profit characteristics.

To investigate this, we repeat the analysis for the smallest and largest quarter of banks in terms of GTA in each year, estimating separate cost and profit functions and using the average level of business conditions specific to each subsample. In 1997, the smallest quarter of banks ranges from about $1 million to about $33 million in assets and the largest quarter ranges from about $134 million to about $278 billion (all in 1994 dollars). The smallest quarter are unambiguously small, but the large quarter encompasses some medium-sized banks because there are not enough large banks alone to accommodate the number of parameters in the models.

The results, shown in Table IV, suggest that our main results and conclusions about cost versus profit productivity are supported, but there are some interesting differences between small and large banks.[18] Cost productivity worsens over the 1984-1991 and 1991-1997 subintervals for both small and large banks, particularly over the latter interval, supporting the earlier results. Using the other breakout of the time periods, the earlier finding of cost productivity improvement for the 1989-1992 credit crunch subinterval that is overwhelmed by cost productivity deterioration in the 1992-1997 subinterval is replicated only for the large-bank subsample. The earlier finding of profit productivity improvements in all cases is replicated here for all time subintervals.[19] Overall, these data are again consistent with the hypothesis that banks attempted to maximize profits in part by providing additional services that raise revenues more than they raise costs.

[Table IV goes here]

7.4 Potential increase in risk-taking

Another potential explanation of our striking findings is that banks may have taken on additional risks during the boom period of the mid-to-late 1990s by shifting assets from securities to loans, by shifting from safer to riskier loans, by adding off-balance sheet risks, by becoming more highly levered, and so forth. Over the long run, higher risks may result in higher average portfolio returns (although not necessarily higher risk-adjusted returns), and these returns may have been magnified by the unexpected strength of the macroeconomy during this period, which may have provided "lucky" short-term rewards for the additional risk-taking. That is, an unusually high proportion of risky investments may have paid off and raised bank revenues as the strong U.S. economy allowed borrowers and other counterparties to repay an unexpectedly high proportion of their obligations. An increase in bank risk-taking that is not captured in our business condition variables, X_C and X_π, may be measured as a profit productivity increase, even if there is no improvement in actual productivity.[20]

To evaluate this possibility, we first note that our business condition variables, X_C and X_π, do have a number of variables that are correlated with bank risk-taking, so we may have already controlled for much of any change in risk-taking. Both X_C and X_π include equity capital, off-balance activities, and market nonperforming loans to help control for changes in bank leverage risk, off-balance sheet risk, and loan risk, respectively. In addition, X_C includes the quantities of various types of loans and securities to help control for asset shifts and X_π includes the market prices of these assets to help control for the portfolio risks.

Banks might be taking increased risks in other ways not controlled for in our estimations. Nonetheless, virtually all raw-data indicators suggest that risk has not increased. The standard deviation of ROE fell from 0.0780 in 1991 to 0.0363 in 1997, and the standard deviation of ROA fell from 0.00501 to 0.00328 over the same period (Table I). Bank failures (not shown) also dropped precipitously from over 100 per year in the late 1980s and early 1990s to the single digits over the last few years of the sample. Under normal circumstances, these findings might suggest a *decrease* in risk, rather than an increase, although the unexpected strength of the macroeconomy makes it difficult to rule out entirely that such an increase in risk-taking occurred.[21]

7.5 Possible increase in conventional market power

Another possible explanation of the findings is that there was an increase in conventional market power over the 1991-1997 time period, in which product quality remained relatively constant, but banks exercised more market power in setting output prices. This could occur because of the consolidation of the banking industry over this time period. The increase in revenues from the higher output prices could explain the improvement in profit performance.

The worsened cost performance may also result in part from an increase in conventional market power, given that it has been found that banks in more highly concentrated local markets have lower cost efficiency, all else equal, presumably because of reduced managerial effort when competition is lax (Berger and Mester 1997, Berger and Hannan, 1998).

However, there are reasons to suspect this does not explain much of our findings. First, local market concentration changed little on average over time. The average value of HERF, the local deposit market concentration faced by banks, rose only about 1.6% from 0.2324 in 1991 to 0.2361 in 1997 (Table I). Despite the merger wave in banking, most of the activity appears to have been of the market-extension type that joined institutions in different local markets.[22] Second, some recent research also suggests that the banks with persistently high profits in the 1990s are not consistently those with high measured market power (Berger, et al., 2000). In addition, both the X_C and X_π vectors of business conditions control for HERF, market-average input prices faced, and state geographic restrictions on bank competition, and the X_π vector also controls for market-average output prices faced, although prices are imperfectly measured. Thus, changes in market power would most likely be measured as the effects of changes in business conditions, rather than changes in productivity.

To investigate the conventional market power issue further, we run some additional regressions to estimate how changes in market concentration and other changes in competitive conditions in markets affected output prices. As shown in Table V, we regress the prices of each of the four variable outputs–consumer loans, business loans, real estate loans, and securities–on market concentration (HERF), the variables representing state geographic restrictions on competition (UNITB, LIMITB, LIMTBHC, NOINTST, and ACCESS, with STATEB excluded as the base case), and variables measuring the recent merger activity of the bank. The merger variables are dummies for whether the bank engaged in one or more mergers (i.e., absorbed one or more other bank charters) in the current year (MERGE0), prior year (MERGE1), or two years prior (MERGE2). These variables allow for the possibility that mergers result in dynamic changes in competition beyond their effects on HERF and the expected effects of the state geographic restrictions. The models are estimated over 1984-1997 with 163,547 observations. We then calculate the effects of the estimated changes in output prices on variable profits.

[Table V goes here]

The results suggest very little effect of market power increases on our profit performance results. The coefficients on HERF indicate that greater concentration is positively and statistically significantly related to the prices of business loans and securities, positively but insignificantly related to the price of real estate loans, and significantly negatively related to the price of consumer loans. The results also suggest that state geographic restrictions on branching and holding company acquisitions within a state (UNITB, LIMITB, LIMTBHC) had mixed effects on prices. However, the prohibition on interstate banking activities (NOINTST) and the amount of interstate access (ACCESS) have the consistent prediction that a larger potential for interstate competition is related to reduced banking prices, as expected. After controlling for concentration and state restrictions on competition, mergers tend to be significantly negatively related to prices of business loans and securities, significantly positively related to the price of consumer loans, and have a mixed, but overall positive relationship with the price of real estate loans.

Despite the finding that many of these effects are statistically significant, their estimated economic effects on profits are quite small, and the explanatory power of the regressors as a whole is quite low. To illustrate the economic effects of HERF, we calculate the predicted effect of the increase in HERF between 1991 and 1997 on each output price, and then calculate the effects of these changes in output prices on variable profits. The annualized predicted effect of the change in HERF on profits is only 0.021%, quite small compared with the annualized total changes in profit of 12.2%, standard profit productivity of 13.7% and alternative profit productivity of 16.5% over this same period shown in Table III above.[23]

The effects of the merger variables, MERGE0, MERGE1, and MERGE2, were computed similarly. The increase in merger activity between 1991 and 1997 had a negative annualized predicted effect on profits of −0.019% after controlling for the other variables, which is again economically quite small.[24] We also re-estimated the price regressions including variables for whether the bank was acquired by a new top-tier holding company but retained its charter (not shown in Table V, but available from the authors). The measured effects on profits of the mergers and bank holding company purchases together were again negative and very small.[25] Thus, the data suggest that any effect of increases in conventional market power from the consolidation of the banking industry are not likely to explain much of our striking performance results.

7.6 <u>Other dynamic effects associated with bank mergers</u>

Mergers may be also associated with other dynamic changes in productivity and performance beyond the exercise of conventional market power. As examples, merging banks might be more aggressive than others in adopting new technologies, they may bring innovations they have already adopted to the banks they acquire, or they may gain a scale of operations needed to take full advantage of new technologies.

To investigate further whether merger activity may account for some of our findings, we re-estimate our cost and profit functions for each year for a subsample of *merging banks* and *nonmerging banks*. For a given year, the merging-banks subsample includes banks that engaged in mergers (i.e., absorbed one or more other bank charters) in the current year or prior two years (MERGE0, MERGE1, or MERGE2 = 1), and the nonmerging-banks subsample includes all other banks. We evaluate the cost and profit functions at the average level of business conditions for that subsample.[26]

The results shown in TableVI suggest that merging banks may be responsible for much of our main findings for the 1991-1997 subinterval.[27] For this subinterval, merging banks had substantial measured cost productivity deterioration of 15.2% annually and standard and alternative profit productivity improvements of 27.4% and 29.5% annually, respectively. These figures are larger in absolute value than the corresponding productivity changes for the industry as a whole shown in Table III above.[28] The changes in average costs and profits for merging banks over this interval (ΔTOTAL) are also larger than the industry figures. Nonmerging banks also had productivity changes in the same direction over this interval, but these changes were much smaller than those of merging banks and the industry as a whole. If our hypothesis that banks provided additional services or higher service quality that raised costs but also raised revenues by a greater amount is correct, then the data in Table VI suggest that merging banks may have done this more than other banks, perhaps as part of their overall strategies of refocusing their institutions.
[Table VI goes here]

Other explanations of the findings for the merging banks are also possible. Merging banks may have improved their profit performance by shifting their portfolios in ways to take advantage of diversification benefits of the mergers. This would be consistent with studies that found improved profit efficiency for merging banks linked to the diversification of risks (e.g., Akhavein, Berger, and Humphrey, 1997; Berger 1998) and with studies that found greater profit performance for banks that are geographically diversified (Hughes, Lang, Mester, and Moon, 1996, 1999). The literature suggests that improved risk diversification allows banks to shift their asset portfolios from securities to loans and to have more assets and loans per dollar of equity, raising average revenues and profits. However, the merger literature generally does *not* find worsened cost performance from mergers—typically mergers were found to have little effect on bank costs or to improve them slightly (e.g., Berger and Humphrey, 1992; Rhoades, 1998). Thus, the merger efficiency literature suggests why the change in profit productivity for merging banks may be more favorable than the change in cost productivity, but it does by itself explain our observed cost productivity deterioration for merging banks. An additional possibility is that organizational problems/disruptions associated with consolidation activity may be responsible for part of the cost productivity deterioration, but we are unaware of any strong evidence to support this possibility. Finally, increases in conventional market power may also help explain part of the measured profit productivity improvement from mergers, but the evidence given in subsection 7.5 suggests that this is unlikely to explain much of the results.

7.7 Bank entry and exit

As a final check on our results, we also investigate whether our results were primarily due to banks that entered and exited the industry. For example, firms that remained in the industry may not have changed their behavior significantly, but that firms that entered during the 1991-1997 time interval may have had relatively low cost productivity and relatively high profit productivity, or that firms that exited may have had relatively high cost productivity and relatively low profit productivity. There is often significant entry into commercial banking through new charters (de novo entry) or conversions from different types of charter (e.g., savings and loans), and exit through failure or merger. Recall from our definition above that merging banks do not enter or exit in the year of the merger, but the banks whose charters they absorbed are counted as exiting.

We examine this issue by rerunning our estimations using a balanced panel of banks that remained in the sample from 1984 through 1997, and the results are shown at the bottom of Table VI. There are 7,392 banks in the panel, or 52.4% of the 14,095 banks at the beginning of the time interval and 83.5% of the 8,855 banks at the end of the interval. The results for the balanced panel are very similar to those reported in Table III using the full sample. This suggests that our results were primarily driven by banks that remained in the sample over time (including the merging banks), rather than being driven by differences in productivity of banks that entered and exited over the time interval.

8. Conclusions

It seems likely that the banking industry would have had tremendous improvements in productivity in recent years, although conventional measurement does not always show this expected result. Banks have adopted innovations in nonfinancial technologies such as information processing and telecommunications, as well as innovations in financial technologies such as financial engineering and statistical analysis in evaluating risks. In addition, the U.S. banking industry has been substantially deregulated over the last two decades, with geographic and product restrictions greatly relaxed, and the industry has undergone a significant consolidation as thousands of banks have disappeared because of M&As and failures. This paper examines how the performance of the banking industry has been affected by these changes in technology, deregulation, and competition. We measure three components of the changes in performance—changes in managerial best practice, which includes technological change and the extent to which the best-practice firms in the industry adopt it, changes in cross-section inefficiency, i.e., dispersion from this best-practice technology, and changes in business conditions exogenous to the banks. We often focus on productivity change, which combines the changes in best practice and changes in inefficiency.

We perform our analysis using three different optimization concepts—cost minimization, standard profit maximization, and alternative profit maximization. These concepts are based on economic optimization in reaction to market prices and other business conditions, in contrast to some government and research measures of productivity change. In our opinion, profit maximization is superior to cost minimization for studying firm performance because it more completely describes the economic goals of managers and owners, who take revenues into account as well as costs. Prior research and government productivity statistics have almost always neglected revenues, which may have resulted in misleading findings.

Our most striking result is that during 1991-1997, cost productivity in the banking industry worsened while profit productivity improved substantially. The most likely explanation is that revenue-based productivity changes are not accounted for in the cost minimization approach. The data are consistent with the hypothesis that banks tried to maximize profits by raising revenues as well as by reducing costs. Over time, banks have offered wider varieties of financial services and provided additional convenience. These additional services or higher service quality, which are difficult to control for in cost and profit functions, may have raised costs but also raised revenues by more than the cost increases, resulting in measurement of worsened cost productivity, but improved profit productivity. Banks that adopt new technologies early are able to capture the return on these technologies in the form of higher profits. As more and more banks adopt the technology, the returns are competed away. The next innovation produces the same result, with early adopters earning high profits until the innovation diffuses through the industry. As long as the innovation process is ongoing, and different new technologies are introduced at different points in time, the innovating banks can keep earning high profits, although it is not necessarily the same individual banks that are the innovators at different points in time.

Further analysis also suggests that banks involved in merger activity may be responsible for much of our main findings. Merging banks had greater cost productivity deterioration and profit productivity improvements than other banks. Any additional services or higher service quality that raised both costs and revenues on average in the industry may have occurred most often for banks engaging in mergers. Merging banks may also have improved their profit performance on average by shifting their portfolios into investments with higher risk and higher expected return to take advantage of diversification gains from the mergers. This diversification benefit may help explain why the change in profit productivity for merging banks may be more favorable than the change in cost productivity, although it does not by itself explain our observed cost productivity deterioration.

Additional analyses in which the models were rerun for a number of different groupings and in a number of different ways generally supported our findings. Our main results were robust to different choices of the time intervals, the use of separate small-bank and large-bank samples, the exclusion of banks that were not present for the entire time interval, the inclusion of measures of the cost of equity capital, and the use of different "thicknesses" in the thick frontier method for disentangling productivity change into its components.

REFERENCES

AKHAVEIN, J. D., BERGER, A. N., AND HUMPHREY, D. B. (1997). The effects of megamergers on efficiency and prices: evidence from a bank profit function, *Rev. Ind. Organ.* 12, 95-139.

ALAM, I. M. S. (2001). A non-parametric approach for assessing productivity dynamics of large banks, *J. Money, Credit, Banking* 33, 121-139.

BAUER, P. W., BERGER, A. N., AND HUMPHREY, D. B. (1993). Efficiency and productivity growth in U.S. banking, *in* "The Measurement of Productive Efficiency: Techniques and Applications" (H.O. Fried, C.A.K. Lovell, and S.S. Schmidt, Eds.), pp. 386-413. Oxford University Press, Oxford.

BERGER, A. N. (1998). The efficiency effects of bank mergers and acquisition: A preliminary look at the 1990s data, *in* "Bank Mergers & Acquisitions" (Y. Amihud, and G. Miller, Eds.), pp. 79-111. Kluwer Academic, Boston.

BERGER, A. N., BONIME, S. D., COVITZ, D. M., HANCOCK, D. (2000). Why are bank profits so persistent? The roles of product market competition, informational opacity, and regional/macroeconomic shocks, *J. Banking Finance* 24, 1203-1235.

BERGER, A. N., AND HANNAN, T.H. (1998). The efficiency cost of market power in the banking industry: A test of the "quiet life" and related hypotheses, *Rev. Econ. Statist.* 80, 454-65.

BERGER, A. N., AND HUMPHREY, D. B. (1991). The dominance of inefficiencies over scale and product mix economies in banking, *J. Monet. Econ.* 28, 117-148.

BERGER, A. N., AND HUMPHREY, D. B. (1992). Measurement and efficiency issues in commercial banking, *in* "Measurement Issues In the Service Sectors" (Z. Griliches, Ed.), pp. 245-279. National Bureau of Economic Research, University of Chicago Press, Chicago.

BERGER, A. N., AND MESTER, L. J. (1997). Inside the black box: What explains differences in the efficiencies of financial institutions? *J. Banking Finance* 21, 895-947.

BUREAU OF ECONOMIC ANALYSIS (BEA) (December 2000). Gross domestic product by industry and domestic income, detailed data files, www.bea.doc.gov/bea/dn2/gpo.htm.

CORRADO, C., AND SLIFMAN, L. (1999). Decomposition of productivity and unit costs, *Amer. Econ. Rev.* 89, 328-332.

DEAN, E. R., AND KUNZE, K. (1992). Productivity measurement in service industries, *in* "Measurement Issues In the Service Sectors" (Z. Griliches, Ed.), pp. 73-101. National Bureau of Economic Research, University of Chicago Press, Chicago.

DEVANEY, M., AND WEBER, W. (2000). Productivity growth, market structure, and technological change: Evidence from the rural banking sector, *Appl. Finan. Econ.* 10, 587-595.

FIXLER, D., AND HANCOCK, D. (1999). Measuring bank credit services: Large versus small U.S. banks, Federal Reserve System, mimeo.

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

FRAME, W.S., SRINIVASAN, A., AND WOOSLEY, L. (2001). The effect of credit scoring on small business lending, *J. Money, Credit, Banking* 33, 813-825.

HUGHES, J. P., LANG, W., MESTER, L. J., AND MOON, C.-G. (1996). Efficient banking under interstate branching, *J. Money, Credit, Banking* 28, 1043-1071.

HUGHES, J. P., LANG, W., MESTER, L. J., AND MOON, C.-G. (1999). The dollars and sense of bank consolidation, *J. Banking Finance* 23, 291-324.

HUGHES, J. P., LANG, W., MESTER, L. J., MOON, C.-G. (2000). Recovering risky technologies using the almost ideal demand system: An application to U.S. banks, *J. Finan. Services Res.* 18, 5-27.

HUGHES, J. P., AND MESTER, L. J. (1993). A quality and risk-adjusted cost function for banks: Evidence on the "too-big-to-fail" doctrine, *J. Productiv. Anal.* 4, 293-315.

HUGHES, J. P., AND MESTER, L. J. (1998). Bank capitalization and cost: Evidence of scale economies in risk management and signaling, *Rev. Econ. Statist.* 80, 314-325.

HUMPHREY, D. B. (1993). Cost and technical change: Effects from bank deregulation, *J. Productiv. Anal.* 4, 9-34.

HUMPHREY, D. B., AND PULLEY, L. B. (1997). Banks' responses to deregulation: Profits, technology, and efficiency, *J. Money, Credit, Banking* 29, 73-93.

KUNZE, K., JABLONSKI, M., AND SIELING, M. (1998), Measuring output and labor productivity of commercial banks (SIC 602): A transactions-based approach, Office of Productivity and Technology, Bureau of Labor Statistics, U.S. Department of Labor.

MESTER, L. J. (1997). What's the point of credit scoring? *Fed. Reserve Bank Bus. Rev. Phila.*, 3-16.

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" (Z. Griliches, Ed.), pp. 25-68. National Bureau of Economic Research, University of Chicago Press, Chicago.

RHOADES, S. A. (1998). The efficiency effects of bank mergers: An overview of case studies of nine mergers, *J. Banking Finance* 22, 273-291.

SEALEY, C., AND LINDLEY, J. (1977). Inputs, outputs, and a theory of production and cost at depository financial institutions, *J. Finance* 32, 1251-1266.

STIROH, K.J. (2000). How did bank holding companies prosper in the 1990s? *J. Banking Finance* 24, 1703-1745.

WHEELOCK, D. C., AND WILSON, P. W. (1999). Technical progress, inefficiency, and productivity change in U.S. banking, 1984-1993, *J. Money, Credit, Banking* 31, 212-234.

YUSKAVAGE, R. E. (1996). Improved estimates of gross product by industry, 1959-94, *Surv. Curr. Bus.* 76, 133-155.

TABLE I

Raw Data Measures of Profits, Cost, Risk, and Market Concentration Unweighted Means for the U.S. Banking Industry, 1984-1997

Year	Number of Banks	Return on Equity	Return on GTA	Costs/ Equity	Costs/ GTA	Revenues/ Costs	Nonperforming Loans/Total Loans	Std. Dev. of ROE	Std. Dev. of ROA	HERF
1984	14,095	0.0649	0.00688	1.248	0.0959	1.163	0.0518	NA	NA	0.2248
1985	14,027	0.0430	0.00570	1.195	0.0904	1.175	0.0576	NA	NA	0.2234
1986	13,734	0.0111	0.00424	1.153	0.0827	1.166	0.0549	0.0949	0.00544	0.2235
1987	13,237	0.0198	0.00480	1.085	0.0784	1.178	0.0473	0.1009	0.00585	0.2258
1988	12,639	0.0507	0.00649	1.083	0.0794	1.182	0.0415	0.0963	0.00584	0.2321
1989	12,295	0.0705	0.00756	1.131	0.0851	1.181	0.0409	0.0943	0.00584	0.2333
1990	11,968	0.0568	0.00656	1.130	0.0844	1.172	0.0421	0.0903	0.00550	0.2332
1991	11,623	0.0714	0.00715	1.037	0.0802	1.183	0.0409	0.0780	0.00501	0.2324
1992	11,238	0.1049	0.00993	0.832	0.0693	1.257	0.0334	0.0690	0.00474	0.2348
1993	10,785	0.1160	0.01101	0.698	0.0608	1.290	0.0285	0.0621	0.00469	0.2376
1994	10,289	0.1142	0.01072	0.699	0.0606	1.298	0.0251	0.0552	0.00434	0.2363
1995	9,747	0.1135	0.01096	0.700	0.0646	1.285	0.0265	0.0467	0.00391	0.2342
1996	9,303	0.1151	0.01107	0.696	0.0644	1.296	0.0275	0.0389	0.00341	0.2369
1997	8,855	0.1149	0.01134	0.685	0.0659	1.294	0.0257	0.0363	0.00328	0.2361

Source: Reports of Condition and Income (Call Reports). The Call Report definitions given below pertain to the 1997 report for banks with domestic and foreign offices (Form FFIEC 031). Comparable definitions were used for other banks and other years.

Notes: GTA refers to gross total assets, which equals total assets (RCFD2170) plus loan and lease loss reserves (RCFD3123) and allocated transfer risk reserve (a reserve for certain foreign loans) (RCFD3128).
Return on equity is net income/total equity capital (= RIAD4340/RCFD3210).
Return on assets is net income/GTA (=RIAD4340/GTA).
Costs are total operating expenses = interest expenses + noninterest expenses (=RIAD4073+RIAD4093).
Revenues are operating income (= RIAD4000).
Nonperforming loans are loans that are past due at least 90 days or are on nonaccrual basis = total loans and lease financing receivables past due 30-89 days and still accruing, past due 90 days or more and still accruing, nonaccruing (= RCFD1406+RCFD1407+RCFD1403).
Std. Dev. of ROE is the standard deviation of ROE over the past five, four, or three years (the longest for which data are available) of the bank's annual return on equity; similarly for Std. Dev. of ROA.
HERF is the weighted-average Herfindahl index of local deposit market concentration across the bank's markets, where each weight is the bank's deposit share in the market. Let d^{jk} = bank j's deposits in market k, then HERF for bank j =

$$\text{HERF}^{j} \equiv \sum_{\text{market}=k} \text{SHARE}^{jk} \times \text{MKTHERF}^{k}, \quad \text{where} \quad \text{SHARE}^{jk} = \frac{d^{jk}}{\sum_{\text{bank}=i} d^{ik}} \quad \text{and} \quad \text{MKTHERF}^{k} = \sum_{\text{bank}=i} \left(\text{SHARE}^{ik}\right)^2 .$$

TABLE II

Variables Used in the Cost, Standard Profit, and Alternative Profit Functions
Means and Standard Deviations for 1984, 1991, and 1997

All financial variables measured in 1000's of constant 1994 dollars using the GDP implicit price deflator.
Prices of financial assets and liabilities are measured as interest rates.
Flow data are for the entire year, stock data are end-of-year values.
See the data appendix for Call Report definitions of variables.

Symbol	Definition	1984 14,095 obs. Mean (Std. Dev.)	1991 11,623 obs. Mean (Std. Dev.)	1997 8,855 obs. Mean (Std. Dev.)
	Dependent Variables			
C	Variable operating plus interest costs, includes costs of purchased funds, core deposits, and labor.	19,191 (244,441)	18,786 (217,483)	22,030 (280,864)
π	Variable profits, includes revenues from loans (minus provisions for loan and lease losses and allocated transfer risk) and revenues from cash and securities, less variable costs.	3,606 (33,750)	4,852 (32,776)	9,662 (99,310)
	Variable Input Prices			
w_1	Market-average price of purchased funds (time deposits over $100,000, foreign deposits, federal funds purchased, demand notes issued to the U.S. Treasury, trading liabilities, other borrowed money, mortgage indebtedness and obligations under capitalized leases, and subordinated notes and debentures).	0.0757 (0.0156)	0.0549 (0.00987)	0.0432 (0.0051)
w_2	Market-average price of core deposits (domestic transactions accounts, time deposits under $100,000, and savings deposits).	0.0682 (0.0152)	0.0428 (0.00822)	0.0231 (0.00666)
w_3	Market-average price of labor (1000's of constant dollars per employee).	30.7 (4.81)	32.5 (6.06)	36.2 (7.26)
	Variable Output Quantities (Cost and Alternative Profit Functions Only)			
y_1	Consumer loans (loans to individuals, bank's domestic offices).	24,885 (189,396)	33,368 (264,761)	52,145 (498,282)
y_2	Business loans (all loans other than consumer and real estate loans).	86,230 (1,392,296)	78,779 (1,178,579)	125,006 (2,151,207)
y_3	Real estate loans (bank's domestic offices).	35,982 (321,058)	74,238 (594,825)	120,986 (1,105,862)
y_4	Securities (all non-loan financial assets, i.e., GTA $-$ $(y_1 + y_2 + y_3) - z_2$, where GTA includes all financial and physical assets (without adjustment for loan and lease loss reserves and allocated transfer risk reserve)).	94,025 (950,646)	125,231 (1,109,465)	205,975 (3,154,739)

TABLE II, CON'T.

Symbol	Definition	1984 14,095 obs. Mean (Std. Dev.)	1991 11,623 obs. Mean (Std. Dev.)	1997 8,855 obs. Mean (Std. Dev.)
	Variable Output Prices (Standard Profit Function Only)			
p_1	Market-average price of consumer loans.	0.110 (0.0306)	0.104 (0.0242)	0.0963 (0.0274)
p_2	Market-average price of business loans.	0.147 (0.0428)	0.115 (0.0373)	0.0974 (0.0282)
p_3	Market-average price of real estate loans.	0.0990 (0.0172)	0.0858 (0.0142)	0.0793 (0.00967)
p_4	Market-average price of securities.	0.0755 (0.0112)	0.0612 (0.00945)	0.0447 (0.00763)
	Fixed Netput Quantities			
z_1	Off-balance-sheet items (commitments, letters of credit, derivatives, etc.) measured using Basel Accord conversion factors and risk weights to be risk-equivalent to loans.	23,914 (529,934)	36,018 (787,436)	74,526 (1,601,009)
z_2	Physical capital (premises and other fixed assets).	3,692 (34,467)	4,732 (47,425)	6,785 (75,692)
z_3	Financial equity capital.	14,754 (132,792)	20,883 (147,961)	41,870 (445,268)
	Environmental Variables			
MNPL	Market-average of nonperforming loans (past due at least 90 days or on nonaccrual basis) divided by total loans.	0.0509 (0.0226)	0.0406 (0.0190)	0.0235 (0.00982)
STINC	Real state income growth (decimal) in bank's state.	0.0567 (0.0197)	0.0103 (0.0134)	0.0294 (0.0105)
UNITB	Dummy, equals one for banks in unit banking states.	0.273 (0.446)	0.0 (0.0)	0.0 (0.0)
LIMITB	Dummy, equals one for banks in limited branching states.	0.648 (0.477)	0.418 (0.493)	0.322 (0.467)
STATEB	Dummy, equals one for banks in statewide branching states. Excluded from the regressions as the base case.	0.0785 (0.269)	0.582 (0.493)	0.678 (0.467)
LIMTBHC	Dummy, equals one for banks in 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.	0.621 (0.485)	0.477 (0.499)	0.489 (0.500)
NOINTST	Dummy, equals one for states that do not allow interstate expansions of multibank holding companies.	0.967 (0.178)	0.0583 (0.234)	0.000678 (0.0260)
ACCESS	Proportion of nation's banking assets in states that are allowed to enter the state in which bank is located (equals proportion of national assets in the state for states that do not allow interstate banking).	0.0350 (0.0342)	0.505 (0.312)	0.607 (0.314)

TABLE II, CON'T.

Symbol	**Definition**	**1984 14,095 obs. Mean (Std. Dev.)**	**1991 11,623 obs. Mean (Std. Dev.)**	**1997 8,855 obs. Mean (Std. Dev.)**
HERF	Weighted-average Herfindahl index of local deposit market concentration across the bank's markets, where each weight is the bank's deposit share in the market. (See footnote to Table I.)	0.2248 (0.167)	0.2324 (0.158)	0.2361 (0.146)
INMSA	Dummy, equals one if the bank is in a Metropolitan Statistical Area.	0.449 (0.497)	0.436 (0.496)	0.432 (0.495)
FED	Dummy, equals one if the bank's primary federal regulator is the Federal Reserve.	0.0725 (0.259)	0.0822 (0.275)	0.111 (0.314)
FDIC	Dummy, equals one if the bank's primary federal regulator is the FDIC.	0.593 (0.491)	0.599 (0.490)	0.606 (0.489)
OCC	Dummy, equals one if the bank's primary federal regulator is the OCC. Excluded from the regressions as the base case.	0.335 (0.472)	0.319 (0.466)	0.283 (0.450)

Notes: All stock values are real quantities (in 1994 dollars) as of the December Call Reports. All prices are market averages of flows over the year divided by these stocks. Note that the flow used in constructing w_2 is interest expenditures less service charges on core deposits. The flows used in computing loan prices p_1, p_2, and p_3, are interest income on loans in the category. Since provisions are reported in the Call Report for total loans and not by loan category, we did not adjust income for loan loss provisions when calculating output prices. We did subtract provisions from loan income in calculating variable profit.

A bank's markets are the MSAs and non-MSA counties in which it operates. 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 other bank's share of the total input usage or total output production of all the other banks in the market. A bank's price is then the weighted average of the prices it faces in each of its markets, where the weight is the bank's share of its deposits in that market.

For example, for an input price, we define the price that bank j faces in market M as $\rho^{jM} \equiv \sum_{i=1,i\neq j}^{m} \left(x^i / \sum_{h=1,h\neq j}^{m} x^h \right) \omega^i$, where m is the number of banks operating in market M, x^i is the input use of bank i and ω^i is bank i's input price (which is bank i's expenditures on the input divided by x^i). Bank j's input price is then the weighted average of the prices faced in each of its markets, where the weight is bank j's share of its deposits in that market: i.e., $w^j \equiv \sum_{k=1}^{N} \left(d^{jk} / \sum_{i=1}^{N} d^{ji} \right) \rho^{jk}$, where N is the total number of markets in which bank j operates and d^{kj} is bank j's level of deposits in market k.

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 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 also 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 the text).

TABLE III
Measured Gross Changes in Cost, Standard Profit, and Alternative Profit at U.S. Banks: Total Change, Productivity Change, Business Conditions Change, Best-Practice Frontier Change, and Inefficiency Change

All Banks

1984-97, 1984-91, and 1991-97

No. of banks in the sample in 1984 = 14,095, in 1991 = 11,623, in 1997 = 8,855

	Cost					Standard Profit					Alternative Profit				
	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF
1984-97	1.011	1.090	0.927	1.052	1.036	1.079	1.091	0.989	1.057	1.032	1.079	1.112	0.970	1.093	1.017
1984-91	0.997	1.042	0.957	1.015	1.026	1.043	1.081	0.965	1.058	1.022	1.043	1.113	0.937	1.118	0.996
1991-97	1.027	1.125	0.913	1.093	1.030	1.122	1.137	0.987	1.054	1.078	1.122	1.165	0.963	1.168	0.998

1984-97, 1984-89, 1989-92, and 1992-97

No. of banks in the sample in 1984 = 14,095, in 1989 = 12,295, in 1992 = 11,238, in 1997 = 8,855

	Cost					Standard Profit					Alternative Profit				
	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF
1984-97	1.011	1.090	0.927	1.052	1.036	1.079	1.091	0.989	1.057	1.032	1.079	1.112	0.970	1.093	1.017
1984-89	1.032	1.084	0.953	1.028	1.054	0.994	1.023	0.972	0.964	1.061	0.994	1.035	0.961	1.048	0.987
1989-92	0.871	0.983	0.886	0.928	1.060	1.242	1.294	0.959	1.373	0.943	1.242	1.340	0.927	1.271	1.054
1992-97	1.082	1.133	0.955	1.184	0.956	1.075	1.088	0.988	0.989	1.101	1.075	1.103	0.975	1.153	0.957

Notes: Figures reported are annualized average rates of change over the interval, i.e., the kth roots of the k-period rates of change [e.g., $(\Delta TOTAL_{Ct,t+k})^{1/k}$].

For the cost measures, a number greater than one indicates an adverse shift toward higher costs; a number less than one indicates a favorable shift.
For the profit measures, a number greater than one indicates a favorable shift toward higher profits; a number less than one indicates an adverse shift.

ΔTOTAL $\equiv \{\exp[f_{Ct+k}(\bar{X}_{Ct+k})] \bullet \exp[\overline{\ln u}_{Ct+k}]\} / \{\exp[f_{Ct}(\bar{X}_{Ct})] \bullet \exp[\overline{\ln u}_{Ct}]\}$ for cost;

$\equiv \{\exp[f_{\pi t+k}(\bar{X}_{\pi t+k})] \bullet \exp[\overline{\ln u}_{\pi t+k}] - \theta_{t+k}\} / \{\exp[f_{\pi t}(\bar{X}_{\pi t})] \bullet \exp[\overline{\ln u}_{\pi t}] - \theta_t\}$ for standard profit;

$\equiv \{\exp[f_{a\pi t+k}(\bar{X}_{Ct+k})] \bullet \exp[\overline{\ln u}_{a\pi t+k}] - \theta_{t+k}\} / \{\exp[f_{a\pi t}(\bar{X}_{Ct})] \bullet \exp[\overline{\ln u}_{a\pi t}] - \theta_t\}$ for alternative profit.

ΔPROD ≡ ΔBESTPR • ΔINEFF for cost, standard profit, and alternative profit.

ΔBESTPR $\equiv \{\exp[f_{Ct+k}(\bar{X}_{Ct})] / \exp[f_{Ct}(\bar{X}_{Ct})]\}$ for cost;

$\equiv \{\exp[f_{\pi t+k}(\bar{X}_{\pi t})] - \theta_{t+k}\} / \{\exp[f_{\pi t}(\bar{X}_{\pi t})] - \theta_t\}$ for standard profit;

$\equiv \{\exp[f_{a\pi t+k}(\bar{X}_{Ct})] - \theta_{t+k}\} / \{\exp[f_{a\pi t}(\bar{X}_{Ct})] - \theta_t\}$ for alternative profit.

ΔBUSCOND $\equiv \{\exp[f_{Ct+k}(\bar{X}_{Ct+k})] / \exp[f_{Ct+k}(\bar{X}_{Ct})]\}$ for cost;

$\equiv \{\exp[f_{\pi t+k}(\bar{X}_{\pi t+k})] - \theta_{t+k}\} / \{\exp[f_{\pi t+k}(\bar{X}_{\pi t})] - \theta_{t+k}\}$ for standard profit;

$\equiv \{\exp[f_{a\pi t+k}(\bar{X}_{Ct+k})] - \theta_{t+k}\} / \{\exp[f_{a\pi t+k}(\bar{X}_{Ct})] - \theta_{t+k}\}$ for alternative profit.

ΔINEFF $\equiv \{\exp[\overline{\ln u}_{Ct+k}] / \exp[\overline{\ln u}_{Ct}]\}$ for cost;

$\equiv \langle \{\exp[f_{\pi t+k}(\bar{X}_{\pi t+k})] \bullet \exp[\overline{\ln u}_{\pi t+k}] - \theta_{t+k}\} / \{\exp[f_{\pi t}(\bar{X}_{\pi t})] \bullet \exp[\overline{\ln u}_{\pi t}] - \theta_t\} \rangle /$

$\langle \{\exp[f_{\pi t+k}(\bar{X}_{\pi t+k})] - \theta_{t+k}\} / \{\exp[f_{\pi t}(\bar{X}_{\pi t})] - \theta_t\} \rangle$ for standard profit;

$\equiv \langle \{\exp[f_{a\pi t+k}(\bar{X}_{Ct+k})] \bullet \exp[\overline{\ln u}_{a\pi t+k}] - \theta_{t+k}\} / \{\exp[f_{a\pi t}(\bar{X}_{Ct})] \bullet \exp[\overline{\ln u}_{a\pi t}] - \theta_t\} \rangle /$

$\langle \{\exp[f_{a\pi t+k}(\bar{X}_{Ct+k})] - \theta_{t+k}\} / \{\exp[f_{a\pi t}(\bar{X}_{Ct})] - \theta_t\} \rangle$ for alternative profit.

TABLE IV
Measured Changes in Cost, Standard Profit, and Alternative Profit at Small and Large U.S. Banks

Small Banks: **In each year, the smallest quarter of banks in terms of GTA.**
Size range in 1984 = $1 million to $24 million in assets
Size range in 1989 = $1 million to $26 million in assets
Size range in 1991 = $1 million to $28 million in assets
Size range in 1992 = $1 million to $29 million in assets
Size range in 1997 = $1 million to $33 million in assets

1984-97, 1984-91, and 1991-97
No. of banks in this subsample in 1984 = 3,523, in 1991 = 2,905, in 1997 = 2,213

	Cost					**Standard Profit**					**Alternative Profit**				
	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF
1984-97	0.982	1.045	0.939	1.046	0.999	1.034	1.053	0.982	1.019	1.033	1.034	1.053	0.982	1.035	1.018
1984-91	0.979	1.006	0.974	0.973	1.033	1.039	1.106	0.939	1.053	1.051	1.039	1.126	0.922	1.078	1.045
1991-97	0.985	1.153	0.854	1.111	1.038	1.029	1.028	1.001	1.039	0.989	1.029	1.217	0.845	1.602	0.760

1984-97, 1984-89, 1989-92, and 1992-97
No. of banks in this subsample in 1984 = 3,523, in 1989 = 3,073, in 1992 = 2,809, in 1997 = 2,213

	Cost					**Standard Profit**					**Alternative Profit**				
	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF
1984-97	0.982	1.045	0.939	1.046	0.999	1.034	1.053	0.982	1.019	1.033	1.034	1.053	0.982	1.035	1.018
1984-89	0.977	1.016	0.961	1.008	1.008	1.031	1.038	0.993	1.021	1.017	1.031	1.069	0.965	1.032	0.998
1989-92	0.939	1.038	0.904	0.751	1.382	1.105	1.130	0.978	1.143	0.988	1.105	1.300	0.850	1.352	0.978
1992-97	1.014	1.063	0.954	1.059	1.003	0.997	1.040	0.958	1.049	0.992	0.997	1.108	0.899	1.362	0.824

Large Banks: **In each year, the largest quarter of banks in terms of GTA.**
Size range in 1984 = $ 98 million to $165 billion in assets
Size range in 1989 = $110 million to $191 billion in assets
Size range in 1991 = $113 million to $176 billion in assets
Size range in 1992 = $116 million to $174 billion in assets
Size range in 1997 = $134 million to $278 billion in assets

1984-97, 1984-91, and 1991-97
No. of banks in this subsample in 1984 = 3,523, in 1991 = 2,905, in 1997 = 2,213

	Cost					**Standard Profit**					**Alternative Profit**				
	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF
1984-97	1.014	1.113	0.911	1.104	1.009	1.086	1.107	0.981	1.032	1.072	1.086	1.122	0.967	1.111	1.010
1984-91	0.999	1.051	0.951	1.064	0.987	1.045	1.062	0.985	1.065	0.997	1.045	1.117	0.936	1.085	1.029
1991-97	1.031	1.108	0.931	1.124	0.986	1.135	1.165	0.974	1.044	1.116	1.135	1.232	0.921	1.224	1.007

1984-97, 1984-89, 1989-92, and 1992-97
No. of banks in this subsample in 1984 = 3,523, in 1989 = 3,073, in 1992 = 2,809, in 1997 = 2,213

	Cost					**Standard Profit**					**Alternative Profit**				
	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF
1984-97	1.014	1.113	0.911	1.104	1.009	1.086	1.107	0.981	1.032	1.072	1.086	1.122	0.967	1.111	1.010
1984-89	1.038	1.091	0.952	1.069	1.021	0.985	1.064	0.969	0.991	1.025	0.985	1.030	0.956	1.026	1.004
1989-92	0.867	0.946	0.916	0.784	1.207	1.281	1.319	0.971	1.277	1.033	1.281	1.323	0.968	1.226	1.080
1992-97	1.088	1.139	0.955	1.224	0.931	1.084	1.101	0.984	0.981	1.123	1.084	1.122	0.966	1.125	0.998

Notes: Figures reported are annualized average rates of change over the interval, i.e., the kth roots of the k-period rates of change [e.g., $(\Delta TOTAL_{Ct,t+k})^{1/k}$].

See notes to Table III for definitions of the change variables.

TABLE V

Regressions of Output Prices on Market Concentration, State Geographic Restrictions on Competition, and Mergers

Dep. Var.:	**p_1** Consumer Loans	**p_2** Business Loans	**p_3** Real Estate Loans	**p_4** Securities
Intercept	0.1103** (614.682)	0.1041*** (380.515)	0.08651*** (756.192)	0.05328*** 689.686
HERF	−0.01327** (−33.682)	0.01639*** (27.258)	0.000383 (1.522)	0.008873*** 52.276
UNITB	−0.001539** (−6.129)	0.01473*** (38.436)	−0.001385*** (−8.648)	0.002588*** 23.928
LIMITB	−0.001103** (−7.853)	0.01058*** (49.358)	−0.001527*** (−17.038)	0.005674*** 93.785
LIMTBHC	−0.01104** (−86.172)	0.002828*** (14.462)	0.000523*** (6.397)	0.002390*** 43.309
NOINTST	0.006655** (37.251)	0.01005*** (36.875)	0.005616*** (49.285)	0.008882*** 115.440
ACCESS	−0.009372** (−39.113)	−0.01183*** (−32.355)	−0.007962*** (−52.095)	−0.010540*** −102.124
MERGE0	0.000741* (1.751)	−0.002634*** (−4.079)	−0.000294 (−1.089)	−0.001506*** −8.260
MERGE1	0.001602** (3.654)	−0.001293*** (−1.934)	0.000371 (1.328)	−0.001228* −6.503
MERGE2	0.001675** (3.654)	−0.002266*** (−3.237)	0.000673** (2.301)	−0.000918*** −4.647
Adj. R-squared	0.0827	0.0920	0.0654	0.3753

Columns report multivariate regression coefficients with t-statistics in parentheses. Note that the STATEB dummy variable is omitted as the base case from the multivariate regressions to avoid perfect collinearity.

* Significantly different from zero at the 10% level.
** Significantly different from zero at the 5% level.
*** Significantly different from zero at the 1% level.

TABLE VI
Measured Changes in Cost, Standard Profit, and Alternative Profit at Merging and Nonmerging U.S. Banks and for a Balanced Panel of U.S. Banks

Merging Banks

1984-97, 1984-91, and 1991-97

No. of banks in this subsample in 1984 = 458, in 1991 = 602, in 1997 = 631

	Cost					Standard Profit					Alternative Profit				
	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF
1984-97	1.100	1.229	0.895	1.251	0.982	1.107	1.160	0.955	1.065	1.089	1.107	1.192	0.929	1.139	1.047
1984-91	1.077	1.127	0.956	1.099	1.025	1.019	1.024	0.995	1.109	0.923	1.019	1.112	0.916	1.102	1.009
1991-97	1.129	1.152	0.980	1.269	0.908	1.221	1.274	0.958	1.112	1.146	1.221	1.295	0.943	1.065	1.216

Nonmerging Banks

1984-97, 1984-91, and 1991-97

No. of banks in this subsample in 1984 = 13,637, in 1991 = 11,021, in 1997 = 8,224

	Cost					Standard Profit					Alternative Profit				
	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF
1984-97	0.962	1.032	0.932	0.976	1.058	1.026	1.034	0.992	1.041	0.993	1.026	1.053	0.974	0.917	1.148
1984-91	0.979	1.025	0.955	0.990	1.036	1.033	1.070	0.966	1.051	1.018	1.033	1.103	0.937	1.093	1.008
1991-97	0.943	1.029	0.916	1.001	1.029	1.017	1.029	0.988	1.028	1.001	1.017	1.050	0.968	1.113	0.943

Balanced Panel of Banks

1984-97, 1984-91, and 1991-97

No. of banks in this subsample in each year = 7,392

	Cost					Standard Profit					Alternative Profit				
	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF	ΔTOTAL	ΔPROD	ΔBUSCOND	ΔBESTPR	ΔINEFF
1984-97	1.018	1.092	0.932	1.094	0.998	1.091	1.096	0.995	1.055	1.039	1.091	1.121	0.973	1.092	1.027
1984-91	0.990	1.032	0.959	1.035	0.997	1.047	1.098	0.953	1.043	1.053	1.047	1.121	0.934	1.118	1.002
1991-97	1.050	1.132	0.928	1.129	1.003	1.144	1.159	0.987	1.087	1.066	1.144	1.195	0.957	1.108	1.079

Notes: For a given year, the merging bank subsample includes banks that engaged in mergers and retained their original charters (i.e., absorbed one or more other bank charters) in the current year or prior two years, and the nonmerging bank subsample includes all other banks that were in the total sample used in our main results reported in Table III. The balanced panel of banks includes those banks that remained in the sample from 1984 through 1997.

Figures reported are annualized average rates of change over the interval, i.e., the kth roots of the k-period rates of change [e.g., $(\Delta TOTAL_{Ct,t+k})^{1/k}$].

See notes to Table III for definitions of the change variables.

DATA APPENDIX

Call Report Definitions of Variables Used in the Cost, Standard Profit, and Alternative Profit Functions

All financial variables measured in 1000's of constant 1994 dollars using the GDP implicit price deflator.
Prices of financial assets and liabilities are measured as interest rates.
Flow data are for the entire year, stock data are end-of-year values.
Unless otherwise stated, the Call Report definitions pertain to the 1997 report for banks with domestic and foreign offices (Form FFIEC 031).
Comparable definitions were used for other banks and for other years.

Symbol	Definition	Call Report Definition
	Dependent Variables	
C	Variable operating plus interest costs, includes costs of purchased funds, core deposits, and labor.	[RIAD4174 + RIAD4172 + RIAD4180 + (RIAD4073 − RIAD4174 − RIAD4509 − RIAD4511 − RIADA518 − RIAD4172 − RIAD4180 − RIAD4508)] + [(RIAD4508 − RIAD4080) + (RIAD4509 + RIAD4511 + RIADA518)] + (RIAD4135) = (RIAD4073 − RIAD4080) + (RIAD4135)
π	Variable profits, includes revenues from loans (minus provisions for loan and lease losses and allocated transfer risk) and revenues from cash and securities, less variable costs.	[(RIAD4107 − RIAD4230 − RIAD4243) + (RIAD3521 + RIAD3196 + RIAD4079 − RIAD4080)] − **C**
	Variable Input Prices	
w_1	Market-average price of purchased funds (time deposits over $100,000, foreign deposits, federal funds purchased, demand notes issued to the U.S. Treasury, trading liabilities, other borrowed money, mortgage indebtedness and obligations under capitalized leases, and subordinated notes and debentures).	Bank's price of purchased funds = expense of purchased funds / stock of purchased funds = [RIADA517 + RIAD4172 + RIAD4180 + (RIAD4073 − RIADA517 − RIAD4509 − RIAD4511 − RIADA518 − RIAD4508 − RIAD4172 − RIAD4180)] / [RCON2604 + RCFN2200 + RCFD2800 − (RCFD2170 + RCFD3123 + RCFD3128) − (RCON2200 + RCFN2200 + RCFD2800 + RCFD3210)] Market-average price calculated as explained in the footnote below.
w_2	Market-average price of core deposits (domestic transactions accounts, time deposits under $100,000, and savings deposits).	Banks's price of core deposits = expense of core deposits / stock of core deposits = [(RIAD4508 − RIAD4080) + (RIAD4509 + RIAD4511 + RIADA518)] / [RCON2215 + (RCON2200 − RCON2215 − RCON6645 − RCON6646)] Market-average price calculated as explained in the footnote below.

DATA APPENDIX, CON'T.

Symbol	Definition	Call Report Definition
w_3	Market-average price of labor (1000's of constant dollars per employee).	Bank's price of labor = salaries and employee benefits / number of employees = RIAD4135 / RIAD4150 Market-average price calculated as explained in the footnote below.
	Variable Output Quantities (Cost and Alternative Profit Functions Only)	
y_1	Consumer loans (loans to individuals, bank's domestic offices).	RCON1975
y_2	Business loans (all loans other than consumer and real estate loans).	RCFD2122 + RCFD2123 − RCON1975 − RCON1410
y_3	Real estate loans (bank's domestic offices).	RCON1410
y_4	Securities (all non-loan financial assets)	RCFD2170 + RCFD3123 + RCFD3128 − RCFD2122 − RCFD2145
	Variable Output Prices (Standard Profit Function Only)	
p_1	Market-average price of consumer loans.	Bank's price of consumer loans = interest income on consumer loans less provisions for loan and lease losses and allocated transfer risk allocated to consumer loans / stock of consumer loans = [interest income on consumer loans − ((net charge-offs on consumer loans/total net charge-offs)×(total provisions))] / [stock of consumer loans] = { (RIAD4054 + RIAD4055) − [[(RIAD4656+RIAD4657)−(RIAD4666+RIAD4667)]/ (RIAD4635−RIAD4605)]×(RIAD4230+RIAD4243)] }/ y_1 Market-average price calculated as explained in the footnote below.

DATA APPENDIX, CON'T.

Symbol	Definition	Call Report Definition
$\mathbf{p_2}$	Market-average price of business loans.	Bank's price of business loans = interest income on business loans less provisions for loan and lease losses and allocated transfer risk allocated to business loans / stock of business loans = [interest income on business loans − ((net charge-offs on business loans/total net charge-offs)×(total provisions))] / [stock of business loans] = { (RIAD4054 + RIAD4055) − [[(RIAD4656+RIAD4657)−(RIAD4666+RIAD4667)] / (RIAD4635−RIAD4605)]×(RIAD4230+RIAD4243)] / $\mathbf{y_2}$ = { [(RIAD4010 + RIAD4065) − (RIAD4054 + RIAD4055) − RIAD4011] − [[(RIAD4635−RIAD4656−RIAD4657−RIAD4651 −RIAD4652) − (RIAD4605−RIAD4666−RIAD4667 −RIAD4661−RIAD4662)] / (RIAD4635−RIAD4605)] × (RIAD4230+RIAD4243)] }/ $\mathbf{y_2}$ Market-average price calculated as explained in the footnote below.
$\mathbf{p_3}$	Market-average price of real estate loans.	Bank's price of real estate loans = interest income on real estate loans less provisions for loan and lease losses and allocated transfer risk allocated to real estate loans / stock of real estate loans = [interest income on real estate loans − ((net charge-offs on real estate loans/total net charge-offs)×(total provisions))] / [stock of real estate loans] = { (RIAD4011) − [[(RIAD4651+RIAD4652)−(RIAD4661+RIAD4662)] / (RIAD4635−RIAD4605)]×(RIAD4230+RIAD4243)] } / $\mathbf{y_3}$ Market-average price calculated as explained in the footnote below.
$\mathbf{p_4}$	Market-average price of securities.	Bank's price of securities = interest income on securities / stock of securities = [(RIAD4107 − RIAD4010 − RIAD4065) + (RIAD3521 + RIAD3196 + (RIAD4079 − RIAD4080))] / $\mathbf{y_4}$ Market-average price calculated as explained in the footnote below.

DATA APPENDIX, CON'T.

Symbol	Definition	Call Report Definition
	Fixed Netput Quantities	
z_1	Off-balance-sheet items (commitments, letters of credit, derivatives, etc.) measured using Basel Accord conversion factors and risk weights to be risk-equivalent to loans.	For 1995 (comparable definitions for other years): (RCFD3819 − RCFD3820 − RCFD3429 + RCFD3654 + RCFD3430) + 0.5 (RCFD3650 + RCFD3652) + 0.2 (RCFD3820 + RCFD3428 + RCFD3433) + 0.5 (RCFD3821 − RCFD3822 + RCFD3833 − RCFD3834) + (0.2 × 0.5) (RCFD3822 + RCFD3824) + 0.2 (RCFD3411)
z_2	Physical capital (premises and other fixed assets).	RCFD2145
z_3	Financial equity capital.	RCFD3210
	Environmental Variables	
MNPL	Market-average of nonperforming loans (past due at least 90 days or on nonaccrual basis) divided by total loans.	RCFD1406 + RCFD1407 + RCFD1403 Market-average price calculated as explained in the footnote below.

Notes: All stock values are real quantities (in 1994 dollars) as of the December Call Reports. All prices are market averages of flows over the year divided by these stocks. Note that the flow used in constructing w_2 is interest expenditures less service charges on core deposits. The flows used in computing loan prices p_1, p_2, and p_3, are interest income on loans in the category. Since provisions are reported in the Call Report for total loans and not by loan category, we did not adjust income for loan loss provisions when calculating output prices. We did subtract provisions from loan income in calculating variable profit.

A bank's markets are the MSAs and non-MSA counties in which it operates. 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 other bank's share of the total input usage or total output production of all the other banks in the market. A bank's price is then the weighted average of the prices it faces in each of its markets, where the weight is the bank's share of its deposits in that market.

For example, for an input price, we define the price that bank j faces in market M as $\rho^{jM} \equiv \sum_{i=1,i\neq j}^{m} \left(x^i / \sum_{h=1,h\neq j}^{m} x^h \right) \omega^i$, where m is the number of banks operating in market M, x^i is the input use of bank i and ω^i is bank i's input price (which is bank i's expenditures on the input divided by x^i). Bank j's input price is then the weighted average of the prices faced in each of its markets, where the weight is bank j's share of its deposits in that market: i.e., $w^j \equiv \sum_{k=1}^{N} \left(d^{jk} / \sum_{i=1}^{N} d^{ji} \right) \rho^{jk}$, where N is the total number of markets in which bank j operates and d^{kj} is bank j's level of deposits in market k.

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 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 also 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 the text).

ENDNOTES

1. Throughout this paper, we use gross total assets (GTA)—total assets plus loan and lease loss reserves and allocated transfer risk reserve (a reserve for certain foreign loans)—as a superior measure of bank size to total assets alone, since GTA does not depend on the performance status of the assets. These unweighted mean returns are somewhat lower in recent years than the more often reported asset-weighted mean returns because large banks have tended to have higher profitability than small banks in these years. Performing the analysis on an unweighted basis allows us to extract meaningful amounts of information from banks of all sizes, rather than having the analysis dominated by the largest banks, which tend to show better performance than small banks. For example, mean return on equity rose from 10.6% in 1984 to 13.9% in 1997 at large banks, and from 2.1% to 9.0% at small banks. Nonperforming loans to total loans fell from 4.6% in 1984 to 2.2% in 1997 at large banks and from 5.4% to 2.8% at small banks.

2. See Dean and Kunze (1992), Mohr (1992), and Kunze, Jablonski, and Sieling (1998) for more information on government measures of productivity change in the service sectors and see Fixler and Zieschang (1997) and Fixler and Hancock (1999) for proposed new methods.

3. We review only some of the literature; see also the literature review in Alam (2001).

4. In some cases, the environmental variables are zero-one dummy variables, such as indicator variables for state branching restrictions, rather than being logged continuous variables. In this discussion, we suppress the time and bank subscripts, but it should be understood that all the variables may differ by time and by bank and that the best-practice cost function $f_C(\cdot)$ may vary over time.

5. Note that profit maximization will differ from value maximization to the extent that banks are not risk-neutral; see Hughes, Lang, Mester and Moon (2000).

6. As the scalar θ goes to zero, $\Delta INEFF_\pi$ simplifies to $\exp[\overline{\ln u}_{\pi t+k}] / \exp[\overline{\ln u}_{\pi t}]$, analogous to the cost case.

7. Consistent with prior research (Berger and Mester, 1997; Humphrey and Pulley, 1997), we find that the alternative profit function fits the data better than the standard profit function. For every year, 1984-1997, the alternative profit function yielded a much higher adjusted R^2. On average, the adjusted R^2 was 40% for the alternative function, and 22% for the standard function.

8. Hughes and Mester (1993) provided an empirical test of the role of deposits in bank production.

9. For robustness checks, we tried incorporating estimates of the cost of equity into measured costs and profits—we re-estimated the cost and profit functions using as the cost of equity the purchased funds rate times the quantity of equity or twice this amount. Our empirical results (not shown) were materially unchanged.

10. To reduce approximation problems near the endpoints, we cut 10% off each end of the $[0,2\pi]$ interval, so that the q_k span $[0.1\times2\pi, 0.9\times2\pi]$. The formula for q_k is $0.2\pi - (\mu_k\times a_k) + (\mu_k\times ln(y_k/z_3))$, where $[a_k,b_k]$ is the range of $ln(y_k/z_3)$ over the entire 14-year time interval, and $\mu_k \equiv [(0.9\times2\pi) - (0.1\times2\pi)]/(b_k-a_k)$.

11. The Fourier-flexible form is a global approximation in which the cosine and sine terms can make the approximating function closer to the true path of the data wherever it is most needed. Prior applications found that the Fourier-flexible form fit the data for U.S. banks better than the commonly specified local translog approximation (e.g., Berger and Mester, 1997). Here, the trigonometric terms improved the fit of the alternative profit model substantially (adjusted R^2 increased by 6 percentage points on average), but there was very little change in fit for the cost function. To keep the number of parameters under control, we limit the trigonometric terms to just the terms containing outputs, including all the first- and second-order terms and the own third-order terms in q_k, which gives a total of 36 trigonometric terms and 95 total net free parameters in the model.

12. This is a slight change from the simplified specification of θ in the prior section, which is necessitated by the normalizations by w_3 and z_3 (discussed in the following paragraph). Thus, the dependent variable will be $\ln(1) = 0$ for the firm with the lowest value of (π/w_3z_3) for that year.

13. The homogeneity restriction does not have to be imposed on the alternative profit function, but we impose it to keep the functional forms equivalent.

14. Analogous to the average-practice case, corrections are made to the best-practice cost and profit functions so that they predict average cost and profit correctly.

15. Increases in competitive pressure may alternatively improve performance by encouraging banks to cut other costs. However, this cannot explain the observed deteriorations in productivity and best-practice.

16. In March 1995, Fair, Isaac introduced its "Small Business Scoring Service (SBSS)," a scoring model developed with RMA, a trade association of commercial lenders. The model was built using a sample of more than 5000 small-business loan applications over five years from 17 U.S. banks.

17. Use of the December 31 Call Report data means that the 1989-1992 interval essentially covers the calendar years 1990, 1991, and 1992, which are often used to represent the credit crunch period.

18. We caution that these results should be considered less reliable than those in Table III. The small-bank and large-bank subsamples use only one-quarter of the data for each year, and the thick-frontier subsamples on which the ΔBESTPR and ΔINEFF effects are based each use only one-sixteenth of the data for each year.

19. The ΔBESTPR and ΔINEFF effects are in some cases unrealistic, presumably because of the very small samples used to estimate these effects, so we do not put much faith in these figures.

20. In principle, luck could explain our results even if banks did not increase their risk-taking. However, the increase in revenues is so large that this seems unlikely. It seems more plausible that an increase in risk coupled with good luck could have led to a large increase in revenues.

21. One study examined whether risk increased during the recent boom period by identifying portfolio characteristics associated with high risk (Berger, et al., 2000). The results were mixed, with the only finding of an increase in risk being an increase in off-balance sheet activities, which we control for in X_C and X_π.

22. Local markets, defined as Metropolitan Statistical Areas or non-MSA rural counties, are reasonable market definitions for retail deposits and small business loans, since these products are competed for primarily on a local basis. These market definitions are also usually those used in antitrust analysis.

23. This predicted change was calculated by first multiplying the coefficient of HERF in each of the output price regressions by the change in average HERF over 1991-1997 to get the predicted effect of the increase in concentration on each output price; multiplying each of these predicted price changes by the 1997 level of the corresponding output; and then summing and annualizing to obtain the predicted effect on profits. The predicted effects of the changes in HERF over 1984-1991 and 1984-1997 are 0.044% and 0.029% of profits, respectively.

24. The annualized predicted effect on profits of the change in mergers between 1984 and 1991 is −0.016%, and between 1984 and 1997 is −0.016%.

25. The annualized predicted effect on profits of the change in number of mergers and purchases between 1991 and 1997 is −0.018%, between 1984 and 1991 is −0.058%, and between 1984 and 1997 is −0.033%.

26. Merging banks tend to be larger, have higher average profits, and lower average costs than nonmerging banks. On average, GTA was $1.1 billion in 1984, $1.6 billion in 1991, and $4.2 billion in 1997 for merging

banks, and $212 million in 1984, $247 million in 1991, and $228 million in 1997 for nonmerging banks. Return on equity was 10.63% in 1984, 9.47% in 1991, and 12.99% in 1997 for merging banks, and 6.35% in 1984, 7.02% in 1991, and 11.37% in 1997 for nonmerging banks. Costs/GTA was 9.40% in 1984, 7.82% in 1991, and 6.27% in 1997 for merging banks, and 9.60% in 1984, 8.03% in 1991, and 6.61% in 1997 for nonmerging banks.

27. The results for the merging banks subsample should be considered less reliable than those in Table III, since the numbers of observations on merging banks (458 in 1984, 602 in 1991, and 631 in 1997) constitute only small percentages of the total numbers of observations used in the full sample (3.2% of total observations in 1984, 5.2% in 1991, and 7.1% in 1997). The number of mergers may appear to be quite small until it is remembered that (1) a merger involves two or more banks and we retain only the survivors in our merger subsample; (2) our merger subsample does not include acquisitions in which the bank keeps its charter but becomes part of a new holding company; and (3) small banks were not the survivors of many mergers over our sample period—e.g., in 1997, less than 1% of our small-bank subsample is in the merger subsample, while over 21% of our large-bank subsample is in the merger subsample.

28. In contrast, over the 1984-1991 interval, merging banks had more measured cost productivity deterioration than the industry as a whole, but less measured improvement in profit productivity.