

What Explains the Dramatic Changes in Cost and Profit Performance of the U.S. Banking Industry?

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

February 1999

The views expressed in this paper do not necessarily represent those of the Federal Reserve Bank of Philadelphia, of the Board of Governors of the Federal Reserve System, or of the Federal Reserve System. We thank Dennis Fixler, Diana Hancock, Dave Humphrey, Rick Lang, Mike Mohr, Leonard Nakamura, Jack Triplett, 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, Chicago, IL, the Australian Industry Economic Conference, Canberra, and the Georgia Productivity Workshop, Athens, GA for helpful comments and advice; Seth Bonime and Chris Malloy for excellent research assistance; and Sally Burke for expert editorial assistance.

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

What Explains the Dramatic Changes in Cost and Profit Performance of the U.S. Banking Industry?

Abstract

We investigate the sources of recent changes in the performance of U.S. banks using concepts and techniques borrowed from the cross-section efficiency literature. 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, and that banks provided additional services or higher service quality that raised costs but also raised revenues by more than the cost increases. The results suggest that methods that exclude revenues may be misleading.

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

Keywords: Bank, productivity, efficiency, cost, profit

What Explains the Dramatic Changes in Cost and Profit Performance of the U.S. Banking Industry?

1. Introduction

The cost and profit performance of the U.S. banking industry was relatively poor from the mid-1980s through the early 1990s, but improved dramatically thereafter through 1997, the end of our sample period. We investigate the sources of these changes in performance over time by borrowing some of the recently developed concepts and techniques used in the cross-section efficiency literature. The three main sources of the changes in costs and profits over time that we investigate are: (1) changes in managerial best practice in the industry, (2) changes in cross-section inefficiency or dispersion from best practice, 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 productivity change.

We analyze the sources of changes 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. These concepts have often been successfully applied to the measurement of cross-sectional differences in efficiency; but have much less frequently been applied to measuring changes over time in performance or productivity. 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 here to ensure a comprehensive look at the data.

The banking industry is of research interest and policy relevance because of its importance to the economy as a whole and because of the puzzling trends in performance within this industry in recent times. The mid-1980s to the early 1990s was a period of relatively poor bank performance. 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 has enjoyed substantially improved performance, reaching record profitability for several years in a row.

Some of these performance trends are illustrated in Table 1. The data indicate that the performance of

U.S. banks improved greatly from the beginning of our sample in 1984 to the end of our sample in 1997. Profitability as measured by mean return on equity rose by more than three-quarters from 6.49% in 1984 to 11.48% in 1997. An alternative measure of profitability, mean return on gross total assets or return on GTA, rose by almost two-thirds from 68.8 basis points to 113.4 basis points.¹ The next two columns in Table 1 indicate that cost ratios also improved dramatically over time. Mean costs/equity and costs/GTA fell by almost one-half and one-third, respectively, between 1984 and 1997. The revenues/costs ratio also indicates a substantial rise of over three-quarters in the mean proportional margin by which revenues exceed costs, increasing from 16.3% in 1984 to 29.4% in 1997. The final column in Table 1 also shows 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 5.28% of loans in 1984 to 2.59% in 1997, consistent with improvements in macroeconomic conditions. As shown, the industry has also been consolidating at a rapid rate, with the number of banks declining by 5,240, or more than one-third of the industry total in 13 years, mostly through merger activity. The results below suggest that this consolidation may have played an important role in the observed dramatic changes in bank performance—merging banks appear to have increased costs but more than made up for this by raising revenues even more.

Another fact that is clear from inspection of Table 1 is that most of the performance improvement occurred after 1991. All of the ratios shown improved slightly (although not all monotonically so) from 1984 to 1991, but all improved more—in most cases dramatically more—after 1991. For example, mean return on equity increased by only 10.0% in the seven years from 1984 to 1991 (from 6.49% to 7.14%), or a 1.4% annual rate of increase. In contrast, this ratio increased by 60.8% in the six years from 1991 to 1997 (from 7.14% to 11.48%), or an 8.2% annual rate of increase. Based on these data and other factors, the analysis in this paper

¹Throughout 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 weighted mean returns because large banks have tended to have higher profitability than small banks in these years. We prefer not to use equity or asset weights for estimating means or other measures because the changes in bank performance are generally industry wide. Our main findings generally hold for both large and small banks, although some exceptions are noted below. 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.

will focus 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 three intervals, 1984-1989, 1989-1992, and 1992-1997, essentially breaking out the “credit crunch” interval in the middle of the sample. As discussed below, we conduct a number of other robustness checks, including examining small and large banks separately, analyzing whether mergers explain the results, and assessing the performance effects of entry and exit in the industry.

We choose 1984 as the start of our sample period, in part because prior studies of productivity growth reviewed below found that bank performance and productivity in the years just prior to 1984 were greatly affected by increases in deposit rates occasioned by the deregulation of these rates—primarily a transfer payment from banks to depositors—rather than a “true” change in productivity. In addition, the bank Call Report allows us to include a much richer set of business conditions in the cost and profit models beginning in 1984.

The cost minimization, standard profit maximization, and alternative profit maximization concepts seem particularly appropriate for investigating these changes in performance over time. These concepts are based on economic optimization, taking certain business conditions as given. The cost concept assumes that firms attempt to minimize variable cost, taking input prices, output quantities, and other conditions as given. Both of the profit concepts assume that firms try to maximize variable profits, but the profit concepts differ from one another in terms of which business conditions are taken as given. Standard profit maximization takes output prices as given, whereas alternative profit maximization takes output quantities as given. As discussed in more detail below, profit maximization is superior to cost minimization for most purposes because it is the more accepted economic goal of firms and firms’ owners, who take revenues as well as costs into account when making decisions. To be comprehensive, we use all three concepts in this study. Each has been found in prior cross-section efficiency analysis to add some independent information, and each yielded findings that were remarkably robust to changes in specification (Berger and Mester, 1997).

It is also important for research and policy purposes to decompose productivity change into the changes in the industry’s managerial most efficient best practice versus changes in industry X-inefficiency, or cross-section dispersion from best-practice costs or profits. A best-practice bank is one that has the lowest expected costs or highest expected profits, given the business conditions specified in its cost or profit function, and

reflects the best attempts to use technology and respond to market prices and other business conditions. Given that we use data from virtually the entire U.S. banking industry at each point in time, the estimated best practice reflects the behavior of the best existing banks, but not any “true” efficient point that fully reflects the available technology and optimal responses to market prices and other business conditions.

The changes in managerial best practice may be driven by technological progress, regulatory innovation, or changes in competitive conditions. Technological progress may include both physical advances, such as better information-processing technologies, or improvements in applied finance, such as derivative contracts. Regulatory innovations include the removal of interstate and intrastate banking regulations, which may make it less costly to raise funds or grow larger. Changes in competitive conditions because of regulatory or market innovations may affect the pressure on best-practice managers to adopt technological improvements or otherwise minimize costs or maximize profits. Inefficiency may be affected by all of these factors as well.²

These arguments imply that it is possible to have worsening performance of the best-practice banks as well as negative productivity change over time. Even if we rule out technological regress, best-practice performance may worsen because of regulatory innovation or changes in competitive conditions. Several studies reviewed below found negative productivity growth for U.S. banks, using cost and nonparametric measures of performance. The goals of our research include determining whether this result continues to hold in our application of the cost minimization concept and whether it also holds in our applications of the profit maximization concepts. If worsening cost performance/productivity change is not accompanied by worsening profit performance/productivity change, this may suggest that applications of cost and perhaps nonparametric concepts that exclude revenue changes may have problems because of unmeasured changes in output quality or failure to adequately capture the goals of banks.

Section 2 reviews prior analysis of bank productivity change from both government statistics and research studies. Section 3 lays out the optimization concepts and how they are applied to decompose changes in performance. Section 4 gives the design of our empirical analysis. Section 5 displays our empirical results, and Section 6 draws conclusions.

²For example, it has been found that banks in more concentrated local markets are significantly less efficient, all else equal, presumably because of reduced managerial effort when competition is lax (Berger and Hannan, 1998).

2. Prior analysis of bank performance change

2.1 Government productivity measures

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 trust accounts (number of these accounts), each weighted by the proportion of employee hours used in each activity. Employee labor hours are used as the denominator of the productivity index, although the BLS also computes an output per employee measure.³ For later comparison to our results and those of other research studies of productivity change, the BLS index for banking productivity per employee hour grew at an annualized rate of 3.25% over 1984-1996, which reflects annualized rates of change of 2.99% and 3.62% over the subintervals 1984-1991 and 1991-1996, respectively (the BLS data for 1997 are not available as of this writing). See Griliches (1992), Dean and Kunze (1992), Mohr (1992), and Kunze, Jablonski, and Sieling (1998) for more information on these and other government measures of productivity change in the service sectors. Fixler and Zieschang (1997) propose a promising financial-firm approach method for measuring bank output, and Fixler and Hancock (1997) and Fixler and Zieschang (1998) propose interesting new methods of measuring the credit services of banks.

2.2 Research studies of productivity change

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

³The U.S. 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).

practice banks and the theoretically best technology available.⁴

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. These three years correspond to pre-, mid-, and post-deregulation of the deposit side of banking. 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 was larger for the smaller banks in the sample. This may reflect the deregulation of deposit rates in the 1980-1984 period, and the fact that smaller banks rely more heavily on deposits than larger banks. Importantly, 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 have outweighed the higher costs to banks. The decline in average costs in the later interval, after most of the effects of deregulation should have been exhausted, may have reflected a decline in market rates that affected deposit rates and rates on purchased funds about equally. When the shifts in the average cost frontiers were adjusted for changes in the 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. Over that period, their estimates ranged from 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 due to high market rates, elimination of deposit rate ceilings, and increased competition from nonbank financial intermediaries, which increased demand for funds and reduced the supply of deposits. Instead of compensating for this increase in input price by closing branches or substituting ATMs, increased competition forced banks to provide more in the way of convenience. Hence, banks increased the number of branches over the 1980s, in addition to paying higher deposit rates and providing the ATM innovation. The increase in deposit rates, increase in nonbank competition, and better

⁴Our discussion reviews only some of the papers in the literature; others include Hunter and Timme (1991), Berg et al. (1992), Elyasiani and Mehdiian (1995), and Hassan, Hunter, and Lozano-Vivas (1998). See also the literature review in Alam (1998).

convenience all made consumers of bank services better off, but because quality of service is difficult to account for in the estimation, the higher quality showed up as a decrease in productivity.

Humphrey (1993) used the same data set to investigate the effect on costs from shifts in the cost function from 1977-1988. Measures were derived in three ways: from a simple time trend; from a time-specific index; and from annual shifts in cross-section cost functions. All three methods yielded similar estimates, with shifts in the cost function implying cost increases averaging 0.8% to 1.4% per year, and small banks (with assets of \$100-\$200 million) experiencing larger increases on average than large banks. Humphrey attributed the decline in cost productivity to deregulation of deposit rates. As support for this hypothesis, he found that in the pre-deregulation period (1977-1980), productivity increased, while during deregulation (1981-1982), productivity declined substantially, and in the post-deregulation period (1983-1988), it showed little change.

Again using the same data set, 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 contemporaneous 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, with assets between \$100 million and \$500 million, the authors found that 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.

In contrast to these several cost minimization studies and one alternative profit maximization study, several research efforts used linear programming methods to measure changes in productivity or performance. 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 (1996) investigated whether the market structure of rural banking markets affected the banks' 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. The authors 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 (1996) also used the linear programming approach to investigate bank productivity growth, decomposing the change in productivity into its change in efficiency and frontier shift components. They found that larger banks (assets over \$300 million) experienced productivity growth between 1984-1993, while smaller banks experienced a decline. Average inefficiency remained high in the industry, since banks were not able to adapt quickly to changes in technology, regulations, and competitive conditions.

Similarly, Alam (1998) used linear programming techniques to investigate productivity change in banking using a balanced panel of 166 banks with greater than \$500 million in assets and uninterrupted data from 1980 to 1989. As in Wheelock and Wilson, productivity change was decomposed into its two components: changes in efficiency and shifts in the frontier. Bootstrapping methods were used to determine confidence intervals for the productivity measure and its components. The findings were that productivity surged between 1983 and 1984, regressed 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.

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

3.1 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:

$$\ln C = f_C(X_C) + \ln u_C + \ln \varepsilon_C. \quad (1)$$

The variable $\ln C$ measures log of variable costs (including both operating and interest expenses); $f_C(\cdot)$ is the best-practice (log) cost function; $X_C \equiv (\ln w, \ln y, \ln z, \ln v)$ is the set of logged exogenous “business conditions” that affect costs, specifically variable input prices ($\ln w$), variable output quantities ($\ln y$), fixed netput quantities ($\ln z$),

and environmental variables (lnv).⁵ The lnu_C term denotes an inefficiency factor that is zero for best-practice firms and raises costs above the best-practice level for other firms because of both technical inefficiency (errors in minimizing inputs relative to outputs or maximizing outputs relative to inputs) and allocative inefficiency (errors in responding to relative prices in choosing inputs or outputs). The $ln\epsilon_C$ term is a random error assumed to have zero mean each period. For convenience, we have suppressed 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 may vary over time.

We represent the cost of the industry at time t in terms of our model by the predicted cost of a bank with average business conditions, average inefficiency for the period, and a zero random error. This gives $\exp[f_C(\bar{X}_C)] \cdot \exp[\bar{lnu}_C]$, where \bar{X}_C gives the average values of the business condition regressors at time t and \bar{lnu}_C 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:

$$\Delta TOTAL_{C,t,t+k} \equiv \{\exp[f_{C,t+k}(\bar{X}_{C,t+k})] \cdot \exp[\bar{lnu}_{C,t+k}]\} / \{\exp[f_C(\bar{X}_C)] \cdot \exp[\bar{lnu}_C]\}. \quad (2)$$

As this is a gross change, any figure below 1.00 indicates falling costs, and vice versa for figures over 1.00. 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% over the k -year interval between t and $t+k$.

Importantly, $\Delta TOTAL_{C,t,t+k}$ embodies all the changes in costs between periods t and $t+k$. The gross proportionate change in costs over time is simply the product of the k one-period gross proportionate changes over the intervening periods. That is:

$$\Delta TOTAL_{C,t,t+k} = \Delta TOTAL_{C,t,t+1} \cdot \Delta TOTAL_{C,t+1,t+2} \cdot \dots \cdot \Delta TOTAL_{C,t+k-1,t+k}. \quad (3)$$

This property will also hold for all the other multiperiod measures derived here. We will report in the tables below the annualized average rate of change over the interval, which is the k th root of the k -period rate of change [e.g., $(\Delta TOTAL_{C,t,t+k})^{1/k}$] to make the findings easier to follow.

More important than the total change in cost are the sources of the change. We decompose $\Delta TOTAL_C$

⁵In some cases, the environmental variables are zero-one dummy variables, such as indicator variables for state branching restrictions, rather than being logged continuous variables.

into the gross changes in best practice, inefficiency, and business conditions, each of which hold the other two components constant:⁶

$$\begin{aligned}
 \Delta \text{TOTAL}_{C,t,t+k} &= \{ \exp [f_{C,t+k}(\bar{X}_{C,t})] / \exp [f_{C,t}(\bar{X}_{C,t})] \} \cdot && \text{(Change in best practice)} \\
 &\{ \exp [\ln u_{C,t+k}] / \exp [\ln u_{C,t}] \} \cdot && \text{(Change in inefficiency)} \\
 &\{ \exp [f_{C,t+k}(\bar{X}_{C,t+k})] / \exp [f_{C,t+k}(\bar{X}_{C,t})] \} && \text{(Change in business conditions)} \\
 &\equiv \Delta \text{BESTPR}_{C,t,t+k} \cdot \Delta \text{INEFF}_{C,t,t+k} \cdot \Delta \text{BUSCOND}_{C,t,t+k} . && (4)
 \end{aligned}$$

Thus, the change in costs can be decomposed into three multiplicative terms. The change in best practice, ΔBESTPR_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, ΔINEFF_C and $\Delta \text{BUSCOND}_C$ give the contributions to ΔTOTAL_C from changes in inefficiency and business conditions only, respectively. All three terms are measured as gross changes, so that a finding of 1.05 would indicate best practice, inefficiency, or business conditions had increased or worsened costs by 5% between periods t and $t+k$.

We also compute cost productivity change, which is simply the product of the change in best practice and the change in inefficiency:

$$\begin{aligned}
 \Delta \text{PROD}_{C,t,t+k} &\equiv \Delta \text{BESTPR}_{C,t,t+k} \cdot \Delta \text{INEFF}_{C,t,t+k} \\
 &= \{ \exp [f_{C,t+k}(\bar{X}_{C,t})] / \exp [f_{C,t}(\bar{X}_{C,t})] \} \cdot \{ \exp [\ln u_{C,t+k}] / \exp [\ln u_{C,t}] \} . && (5)
 \end{aligned}$$

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 for several reasons. First, productivity change is an established and informative concept. As shown in Section 2 above, productivity change is used in prior research, which separates the changes in the behavior of firms from changes in the business conditions they face. One of the goals of this paper is to determine if and why productivity change in banking based on costs or nonparametric measures is sometimes negative.

Second, in our opinion, cost productivity change represents an improvement over government

⁶To obtain equation (4), we multiply the numerator and denominator on the right-hand-side of equation (2) by $\exp [f_{C,t+k}(\bar{X}_{C,t})]$ — what the best-practice cost at time $t+k$ would be if business conditions had remained unchanged since time t —and rearrange terms.

productivity change statistics. As discussed, the government measures use the change in a single output—gross product originating or a weighted sum of bank transactions—divided by the quantity of a single input measure, employee labor hours. Assuming that costs are proportionate to these hours and that the effect of business conditions as a whole are proportional to the effect of the single measured output, the government productivity change measures should give qualitatively similar results to those given by ΔPROD_C . Thus, if these conditions hold, a ΔPROD_C of 1.05 should roughly correspond to a government productivity change of -5%.

However, ΔPROD_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 all of these factors, so that an improvement or deterioration in costs that is not due to any decision or managerial skill of the bank is not attributed to a change in productivity.

It is also important in our opinion to include 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 labor hour may overstate the change in output per total labor hour worked by employees and nonemployees. As of 1992 (a benchmark year for the BEA), labor expenses accounted for only 35.8% of the BEA's gross product originating in the banking sector. 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 the labor may be employed elsewhere in the holding company (Berger, Kashyap, and Scalise, 1995). 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.

A third, more pragmatic reason to measure productivity change in addition to the change in best practice and change in inefficiency is that ΔPROD_C is relatively easy to estimate, whereas dividing it into ΔBESTPR_C versus ΔINEFF_C is more difficult and is likely to involve more estimation error. As explained further below, we use a simple average-practice cost function to separate productivity change ΔPROD_C from the change in business conditions ΔBESTPR_C , but further decomposing ΔPROD_C into ΔBESTPR_C versus ΔINEFF_C requires the additional use of the thick frontier or other method to measure changes in the best-practice frontier.

3.2 Standard profit maximization

The two profit maximization concepts assume that firms maximize variable profits, again subject to a number of exogenously given business conditions. Standard and alternative profit maximization differ from one another only in terms of the specification of the business conditions. In our opinion, profit maximization is superior to cost minimization for the study of firm performance because it more completely describes the economic goals of firms and their 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 generally will not be counted as an improvement and may be counted as a deterioration under cost minimization. One firm might do a slightly better job at minimizing its costs than another firm, but it might make less profits than the other firm because it does a much worse job at choosing its output mix. The nonparametric methods and government productivity statistics described above may also neglect beneficial effects of revenue gains.⁷ Hence, focusing on cost minimization alone would give an incomplete picture of firm performance.

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:

$$\ln(\pi + \theta) = f_{\pi}(X_{\pi}) + \ln u_{\pi} + \ln \varepsilon_{\pi} , \quad (6)$$

⁷Note that profit maximization will differ from value maximization to the extent that banks are not risk-neutral; see Hughes, Lang, Mester and Moon (1997).

where π is the variable profits of the firm, which includes all the interest and fee income earned on the variable outputs minus variable costs C , 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 does vary over time). The vector of business conditions $X_\pi \equiv (lnw, lnp, lnz, lnv)$ is the same as X_C , except logged output prices lnp replace logged output quantities lny . Analogous to the case of the cost function, $f_\pi(\cdot)$ is the best-practice profit function, lnu_π 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. We treat θ as a time-varying parameter that is part of the best-practice profit function along with the other parameters in $f_\pi(\cdot)$.

The decomposition of the change in profit over time is quite 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 $\exp[f_\pi(\bar{X}_\pi)] \cdot \exp[\bar{lnu}_\pi] - \theta_t$, and the total gross change in profit between periods t and $t+k$ is given by:

$$\Delta TOTAL_{\pi,t,t+k} = \{ \exp[f_{\pi,t+k}(\bar{X}_{\pi,t+k})] \cdot \exp[\bar{lnu}_{\pi,t+k}] - \theta_{t+k} \} / \{ \exp[f_\pi(\bar{X}_\pi)] \cdot \exp[\bar{lnu}_\pi] - \theta_t \}. \quad (7)$$

Here, a figure above 1.00 indicates an improvement in profits, so that a figure of 1.05 would indicate that profits have increased or improved by 5% over the k -year interval between t and $t+k$.

The components of $\Delta TOTAL_{\pi,t,t+k}$ are decomposed as:

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

$$\Delta INEFF_{\pi,t,t+k} = \langle \{ \exp[f_{\pi,t+k}(\bar{X}_{\pi,t+k})] \cdot \exp[\bar{lnu}_{\pi,t+k}] - \theta_{t+k} \} / \{ \exp[f_\pi(\bar{X}_\pi)] \cdot \exp[\bar{lnu}_\pi] - \theta_t \} \rangle / \langle \{ \exp[f_{\pi,t+k}(\bar{X}_{\pi,t+k})] - \theta_{t+k} \} / \{ \exp[f_\pi(\bar{X}_\pi)] - \theta_t \} \rangle$$

$$\Delta BUSCOND_{\pi,t,t+k} = \{ \exp[f_{\pi,t+k}(\bar{X}_{\pi,t+k})] - \theta_{t+k} \} / \{ \exp[f_{\pi,t+k}(\bar{X}_\pi)] - \theta_{t+k} \}. \quad (8)$$

Analogous to the case for the cost decomposition, $\Delta TOTAL_{\pi,t,t+k} = \Delta BESTPR_{\pi,t,t+k} \cdot \Delta INEFF_{\pi,t,t+k} \cdot \Delta BUSCOND_{\pi,t,t+k}$. Note that the change in inefficiency $\Delta INEFF_\pi$ is simply the ratio of the gross change in profit inclusive of the change in inefficiency to what the gross change would be if there were no inefficiency. As the scalar θ goes to zero, this formula simplifies to the ratio of the inefficiency terms, $\exp[\bar{lnu}_{\pi,t+k}] / \exp[\bar{lnu}_\pi]$,

analogous to the cost case. Also consistent with the cost minimization case, standard profit productivity ΔPROD_π is the product of ΔBESTPR_π and ΔINEFF_π .

3.3 Alternative profit maximization

The alternative profit maximization concept has the same objective as the standard profit maximization concept, but specifies the same set of business conditions be taken as given as the cost minimization concept. The alternative profit function is given by:

$$\ln(\pi + \theta) = f_{a\pi}(X_C) + \ln u_{a\pi} + \ln \varepsilon_{a\pi}. \quad (9)$$

That is, the logged output quantities $\ln y$ are specified in the X vector, rather than logged output prices $\ln p$. 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 , which is the same as the gross change in average variable profits, but the decompositions into the various components will differ. The formulas follow the standard profit maximization case except for the use of the slightly different business conditions.

It is not expected that firms actually take their outputs as given and maximize profits, as the alternative profit specification literally implies. There would be no reason to 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. These conditions were laid out in terms of cross-section differences applicable to efficiency measurement, but we apply them here equally well 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, the 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 may not, because they take into account the extra revenues that cover these costs. Standard profit measures, by virtue of including revenues, may also be better able to handle changes in output quality than cost measures or other measures that do not take revenues into account. Second, the variable outputs may not be completely variable, as is 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 assumption of exogenous prices in the standard profit maximization concept. 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 ΔBESTPR_π , ΔINEFF_π , and $\Delta\text{BUSCOND}_\pi$. The alternative profit function may be more accurately measured, or at least provide an alternative measurement of the components of the change in profitability to check robustness.⁸ 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 to the data.

4. Methodological design

Our data set includes annual information on virtually all U.S. commercial banks from 1984 through 1997, although we primarily focus on data from 1984, 1991, and 1997, as indicated earlier. Because of consolidation in the industry, the number of observations declines from 14,095 in 1984 to 11,623 in 1991 to 8,855 in 1997, as shown in Table 1 above. Most of the data are from the Reports of Income and Condition (Call Reports).

4.1. Variables

Table 2 gives the definitions of all the variables specified in the cost and profit functions, as well as their sample means and standard deviations for 1984, 1991, and 1997. Although the continuous variables are

⁸Consistent 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 each of the 14 years, 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. A better fit does not prove in any rigorous sense that the right-hand-side variables are better measured, but it is consistent with the hypothesis that the output variables as a group may be measured with less noise.

generally expressed in natural logs in the cost and profit functions, we show the means and standard deviations of the levels to be more informative. Recall that the cost and alternative profit specifications are identical except for the dependent variable, and the exogenous variables or business conditions differ for the standard profit function in that variable output prices are used in place of variable output quantities.

In choosing which financial accounts to specify as outputs versus inputs, we use the “asset approach” or “intermediation approach” of Sealey and Lindley (1977). Under the asset approach, banks are considered as financial intermediaries between those that provide funds to the bank and those that receive funds from the bank. All liabilities (core deposits and purchased funds) and financial equity capital provide funds and are considered to be inputs that generate costs. All assets (loans and securities) use bank funds and are considered outputs that generate revenues. Physical inputs (labor and premises) are specified as inputs that generate costs. The financial inputs are sometimes paid for with depositor services instead of with interest payments, but this is incidental under the asset approach, since the funds raised would be worthless without the assets in which they are invested.

Other approaches to choosing the specification are the value added and user cost approaches, which typically specify depositor services as an important final output. Since these services are difficult to measure, they are usually assumed to be proportionate to the quantity of deposits, which is then specified as an output in the empirical model. In our opinion, the asset approach—which treats deposits as an input—is most compatible with the profit maximization concepts discussed earlier. This is because deposits and other liabilities by themselves generate negative cash flows and reduce profits, whereas loans and other assets generate positive cash flows and profits.⁹

As a practical matter, it would also be difficult to specify a positive output price for depositor services under the other approaches, since most of these services are not explicitly priced. The cost function literature has generally yielded qualitatively similar results whether deposits are treated as an input or output (e.g., Mester, 1992). However, no one to our knowledge has specified deposits or depositor services as variable outputs in

⁹See Triplett (1998) for extensive discussion of these and other measurement issues regarding bank outputs and inputs.

a profit function, perhaps in part because of the difficulty of finding a positive output price.¹⁰

For all of 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 detail in the notes to Table 2, only data from other firms in the bank's local markets are used to construct these market-average prices. Note also that the prices take into account banks' fee income and expense as well as interest income and expense. One advantage of the market-average prices faced is that they 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 having a high lnu_c , low lnu_π , and low lnu_{act} , reflecting its inefficiency at minimizing costs or maximizing profits given its market-average deposit rate faced in the lnw vector and other exogenous business conditions in the entire X vector. A third advantage of market-average prices is that they are likely to average out some of the computational errors in measuring prices paid or received by individual banks.

The variable inputs for which prices lnw are specified are purchased funds, core deposits, and labor. The variable outputs lny 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 lnz . For the off-balance-sheet items, we use the Basle 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 outputs 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 a price for these durable inputs. As noted above, financial equity capital is an input under the asset approach, and we treat it as fixed here, in part because it is difficult to change quickly and in part because its

¹⁰One cost study derived an empirical test for determining whether deposits are inputs or outputs and found that deposits should be treated as an input (Hughes and Mester, 1993). There is also some evidence that information gathered from depositors is used as an input in the loan process (Allen, Saunders, and Udell, 1991; Nakamura, 1993; Mester, Nakamura, and Renault, 1998).

price, the risk-adjusted expected return on equity, is difficult to measure. In addition, banks must meet minimum financial capital requirements, including equity, 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 keep risks under control (Hughes and Mester, 1993, 1998; Hughes, Lang, Mester, and Moon, 1997). For robustness checks, we tried incorporating estimates of the cost of equity into measured costs and profits (not shown), and our empirical results were materially unchanged.¹¹

Among the environmental variables lnv we include the log of the market-average nonperforming loans to total loans ratio, $lnMNPL$, and $\frac{1}{2}(lnMNPL)^2$, since dealing with exogenous loan problems raises costs and lowers profitability (Berger and DeYoung, 1997). As is the case for prices, we use the market average rather than each individual bank's nonperforming loan ratio, since the market average captures the exogenous conditions in markets that affect loan performance and is less likely to reflect decisions made by the bank's own managers. Market conditions are also accounted for by state income growth ($STINC$, $\frac{1}{2}STINC^2$). We also 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 ($LIMITBHC$); whether out-of-state holding company expansion is prohibited ($NOINTST$); and the proportion of the U.S. banking industry's assets held in states allowed to enter the bank's own state ($ACCESS$, $\frac{1}{2}ACCESS^2$). Other controls 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).

4.2 Functional form

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

¹¹The required rate of return on equity is likely to be somewhat greater than the rate faced on purchased funds. We tried reestimating the cost and profit functions using as the cost of equity either the purchased funds rate times the quantity of equity or twice this amount. Both reestimations yielded results very similar to those obtained with our main specification reported in Table 3.

$$\begin{aligned}
\ln(C/w_3z_3) = & \alpha + \sum_{i=1}^2 \beta_i \ln(w_i/w_3) + 1/2 \sum_{i=1}^2 \sum_{j=1}^2 \beta_{ij} \ln(w_i/w_3) \ln(w_j/w_3) + \sum_{k=1}^4 \gamma_k \ln(y_k/z_3) \\
& + 1/2 \sum_{k=1}^4 \sum_{m=1}^4 \gamma_{km} \ln(y_k/z_3) \ln(y_m/z_3) + \sum_{r=1}^2 \delta_r \ln(z_r/z_3) \\
& + 1/2 \sum_{r=1}^2 \sum_{s=1}^2 \delta_{rs} \ln(z_r/z_3) \ln(z_s/z_3) + \sum_{i=1}^2 \sum_{k=1}^4 \eta_{ik} \ln(w_i/w_3) \ln(y_k/z_3) \\
& + \sum_{i=1}^2 \sum_{r=1}^2 \rho_{ir} \ln(w_i/w_3) \ln(z_r/z_3) + \sum_{k=1}^4 \sum_{r=1}^2 \tau_{kr} \ln(y_k/z_3) \ln(z_r/z_3) \\
& + \sum_{k=1}^4 [\varphi_k \cos(q_k) + \omega_k \sin(q_k)] + \sum_{k=1}^4 \sum_{m=k}^4 [\varphi_{km} \cos(q_k+q_m) + \omega_{km} \sin(q_k+q_m)] \\
& + \sum_{k=1}^4 [\varphi_{kkk} \cos(q_k+q_k+q_k) + \omega_{kkk} \sin(q_k+q_k+q_k)] + \sum_{n=1}^{14} \xi_n \ln v_n + \ln u_C + \ln \varepsilon_C, \quad (10)
\end{aligned}$$

which is estimated separately for each year's data, allowing all the parameters to vary over time. The variables (y_k/z_3) , (z_r/z_3) , and MNPL (included in the environmental variables $\ln v$) have 1 added before logging for every firm in order to avoid taking the natural 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).¹² 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 exclude factor share equations, which embody restrictions imposed by Shephard's Lemma (or Hotelling's Lemma for the standard profit function), because these would impose the undesirable assumption of perfect allocative efficiency (no errors in responding to relative prices in choosing inputs or outputs).

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, the constant $\theta_t \equiv |(\pi/w_3z_3)_t^{\min}| + 1$ is added to every firm's dependent variable in the profit function so that the natural log is taken of a positive number, since the minimum profits are typically negative. This is a slight change from the simplified specification of θ in the prior section, which is necessitated by the normalizations by the last input price w_3 and by the last fixed netput quantity 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. For the alternative profit function, this is the only change in specification (other than relabelling the composite error term as $\ln u_{\text{alt}} + \ln \varepsilon_{\text{alt}}$), 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,

¹²To reduce approximation problems near the endpoints, we cut 10% off each end of the $[0, 2\pi]$ interval, so that the q_k span $[0.1 \times 2\pi, 0.9 \times 2\pi]$. 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 $\mu_k \equiv (0.9 \times 2\pi - 0.1 \times 2\pi)/(b_k - a_k)$.

$\ln(y_k/z_3)$, 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. Normalization by equity makes the dependent variable reasonably equally achievable for all banks. 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.

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. financial institutions better than the commonly specified local translog approximation, although prior studies disagree over whether the efficiency results are very different from those obtained when using the translog specification (McAllister and McManus, 1993; Mitchell and Onvural, 1996; Berger, Leusner, and Mingo, 1997; Berger, Cummins, and Weiss, 1997; Berger and DeYoung 1997; Berger and Mester, 1997). In this study, the inclusion of 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 Fourier 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

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

parameters in the model (after imposing symmetry on the translog portion of the model).

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

We decompose the total changes in costs and profits over time, $\Delta\text{TOTAL}_{C_{t,t+k}}$, $\Delta\text{TOTAL}_{\pi_{t,t+k}}$, and $\Delta\text{TOTAL}_{\text{art},t+k}$ in several steps. First, we estimate simple average-practice cost and profit functions for each year that include all banks without regard to 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 \bar{lnu}). We then use these average-practice functions to separate ΔTOTAL into the productivity change ΔPROD and the change in business conditions $\Delta\text{BUSCOND}$ components.

The changes in $\Delta\text{PROD}_{C_{t,t+k}}$, $\Delta\text{PROD}_{\pi_{t,t+k}}$, and $\Delta\text{PROD}_{\text{art},t+k}$ 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 from the definition of cost productivity change from equation (5) above to give:

$$\begin{aligned}\Delta\text{PROD}_{C_{t,t+k}} &\equiv \Delta\text{BESTPR}_{C_{t,t+k}} \cdot \Delta\text{INEFF}_{C_{t,t+k}} \\ &= \{ \exp [f_{C_{t+k}}(\bar{X}_{C_t})] / \exp [f_{C_t}(\bar{X}_{C_t})] \} \cdot \{ \exp[\bar{lnu}_{C_{t+k}}] / \exp[\bar{lnu}_{C_t}] \} \\ &= \{ \exp [f_{C_{t+k}}(\bar{X}_{C_t})] \cdot \exp[\bar{lnu}_{C_{t+k}}] \} / \{ \exp [f_{C_t}(\bar{X}_{C_t})] \cdot \exp[\bar{lnu}_{C_t}] \} .\end{aligned}\quad (11)$$

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, the cost productivity change is what the gross change in costs would be if business conditions had remained unchanged since time t , inclusive of the change in best practice and the change in inefficiency, both of which are incorporated in the average-practice cost function. We simply use the estimated parameters of average-practice cost functions to evaluate the numerator and denominator of (11). We estimate the changes in costs or profits that are due to changes in business conditions, $\Delta\text{BUSCOND}_{C_{t,t+k}}$, $\Delta\text{BUSCOND}_{\pi_{t,t+k}}$, and $\Delta\text{BUSCOND}_{\text{art},t+k}$ as the changes in costs or profits that remain after accounting for the productivity changes (e.g., $\Delta\text{BUSCOND}_{C_{t,t+k}}$ is estimated from $\Delta\text{TOTAL}_{C_{t,t+k}} / \Delta\text{PROD}_{C_{t,t+k}}$).

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. That is, letting $[\ln(C/w_3z_3)]^{\text{pred}}$ or $[\ln[(\pi/w_3z_3) + \theta]]^{\text{pred}}$ be the predicted values of the dependent variables evaluated at \bar{X}_C or \bar{X}_π , the transformation of these predicted values— $\{\exp[\ln(C/w_3z_3)]^{\text{pred}}\} \cdot w_3 \cdot z_3$ or $\{\exp[\ln[(\pi/w_3z_3) + \theta]]^{\text{pred}} - \theta\} \cdot w_3 \cdot z_3$ —do not precisely equal the average values of C or π for that year. 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 $\Delta\text{BUSCOND}$ correctly multiply to ΔTOTAL .

Finally, we decompose the productivity changes ΔPROD into the change in best practice ΔBESTPR and change in inefficiency ΔINEFF components. This represents a challenge, because there is no consensus as to the best way to estimate the best-practice frontier or even if any of the various methods used are very precise. We use a version of the thick frontier method to measure ΔBESTPR (Berger and Humphrey, 1991). Each year, we divide the banks based on their residuals from estimating the average-practice cost and profit functions, which use the data on all banks operating in the year. Banks with residuals in the “best” 25% in each of 10 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 just this most efficient quarter of banks (i.e., we assume that the residuals represent random error and not differences in inefficiency). These estimated thick frontiers are treated as the best-practice functions $f_C(\cdot)$, $f_\pi(\cdot)$, and $f_{\text{an}}(\cdot)$.¹⁴ The ΔBESTPR are simply measured as the changes in costs or profits due to changes in $f(\cdot)$, holding business conditions constant at period t values. The changes in inefficiency ΔINEFF are simply estimated as the changes in productivity ΔPROD that remain after accounting for the change in best practice ΔBESTPR (e.g., $\Delta\text{INEFF}_{C,t+k}$ is estimated as $\Delta\text{PROD}_{C,t+k} / \Delta\text{BESTPR}_{C,t+k}$). Since there is considerable 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.

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

We also repeat all of these procedures for small and large banks, where small and large are defined as being in the bottom quarter and top quarter, respectively, in terms of GTA in a given year. To give an idea as to these sizes, 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 real 1994 dollars). The smallest quarter are unambiguously small, but the large quarter obviously encompasses some medium-sized banks because there are not enough large banks alone to accommodate the number of parameters in the models. Different functions are estimated to allow for the possibility that small and large banks use different technologies, produce different services, have different efficiency, etc.

5. Empirical results

5.1 Cost and profit changes for all U.S. banks, 1984-1997

The top panel of Table 3 reports the total changes in costs and profits ΔTOTAL_C , ΔTOTAL_π and $\Delta\text{TOTAL}_{\text{net}}$, and the decompositions of these total changes into their ΔPROD , $\Delta\text{BUSCOND}$, ΔBESTPR , ΔINEFF components for all U.S. banks over 1984-1997, and over the two subintervals 1984-1991 and 1991-1997. As discussed above, we correct the average-practice cost and profit functions such that they predict the sample average cost or profit for that year when evaluated at the mean value of business conditions, so that ΔTOTAL_C equals the actual change in average cost and ΔTOTAL_π and $\Delta\text{TOTAL}_{\text{net}}$ both equal the actual change in average profits for the interval. As indicated, the figures reported are annualized gross changes, the k th roots of the k -period rates of change.

The ΔTOTAL_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 1 above, which were scaled by equity or GTA. Here, we include these scale factors in our cost business conditions vector X_C along with the variable input prices, variable output quantities, other fixed netput quantities, and environmental variables. As shown in Table 2 above, 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 ($\Delta\text{PROD}_c > 1$), while the business conditions as a whole reduced costs over both subintervals ($\Delta\text{BUSCOND}_c < 1$). 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%.

If we examine the changes in business conditions more closely, the strong benefits of these changes in lowering costs are not at all surprising. As shown in Table 2, interest rates on liabilities have declined substantially—purchased funds rates faced fell from 7.57% in 1984 to 5.49% in 1991 to 4.32% in 1997, while core deposit rates faced fell from 6.82% to 4.28% to 2.31% over the same time periods. 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 and in the costs of the average banking firm from pure increases in scale. 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 1 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, depending in part on the extent to which consumers received the benefits of the increases in competition (discussed further below).¹⁵

More puzzling are the measured unfavorable shifts in cost productivity of -4.2% annually over 1984-1991 and -12.5% annually over 1991-1997. Using the thick frontier methodology to approximate the best-practice cost function (estimated using the “best” 25% of banks in each of 10 size classes) suggests that both best-practice costs and inefficiency worsened in both subintervals and that an unfavorable shift in best practice

¹⁵Although we are able to identify the likely qualitative effects of some of the changes in business conditions, we were unable to quantify these effects to our satisfaction. We tried to decompose $\Delta\text{BUSCOND}$ into several subcomponents by measuring the effects on costs and profits of changing some of the X variables to the mean values for time t+k and leaving others at the mean values for time t. However, nonlinearities in the interaction terms among the prices, variable outputs, and fixed netputs made this difficult. For example, the $\eta_{ik} \ln(w_i/w_3) \ln(y_k/z_3)$ terms from equation (10) mix two input prices, one variable output, and one fixed netput in a highly nonlinear way. Trying to change some of these variables but not others, combined with other nonlinearities in the cost and profit functions described earlier, made the results of this additional decomposition questionable.

explains most of the worsening 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, since there is considerable 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 3.62% over 1991-1996 reported earlier, although other research studies of productivity change often found productivity deteriorations. As discussed above, we believe our cost productivity change represents an improvement over government productivity change statistics because we control for all of the output quantities, input prices, fixed netput quantities, and environmental conditions specified in the business conditions vector X_c , rather than using the ratio of a single output to a single input. In addition, we include all variable costs, including employee labor hours, which may understate labor input by nonbank employees working elsewhere in the bank holding company, as well as nonbank labor and capital used to raise funds elsewhere and purchased with interest payments.

As discussed above, even if we rule out technological regress, it is possible to have worsening productivity and unfavorable shifts in best practice over time because of regulatory innovation or changes in competitive conditions, and such results have been found in past research (Berger and Humphrey, 1992; Bauer, Berger, and Humphrey, 1993; Humphrey, 1993; Wheelock and Wilson, 1996; Alam, 1998). Previous researchers have often cited the deregulation of deposit interest rates in the early 1980s and the resulting disequilibrium from this regulatory change as important factors in raising costs and worsening performance, but these factors almost surely cannot explain our strong findings of results of worsening productivity/best practice over 1991-1997, which is well after the deregulation of the rates. It is possible that there was an increase in competition due to liberalizations 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, which include measures of state restrictions.¹⁶

Much of the remainder of our analysis will be devoted to trying to explain these puzzling observed deteriorations, as well as checking robustness of our results more generally. A primary candidate for explaining the cost-based deteriorations is that measures based on cost minimization may give an incomplete view of bank performance, since they do not account well for unmeasured changes in output quality and they fail to completely capture the profit maximization goal of banks, which requires that effort be spent to raise revenues as well as reduce costs. A number of other candidates for explaining our findings (time periods chosen, problems with including both small and large banks in the same analysis, merger activity, industry entry and exit) will also be examined below.

The profit figures in the top panel of Table 3 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, rising at an annual rate of 4.3% over 1984-1991, and rising at a very substantial annual rate of 12.2% over 1991-1997. The data suggest that the increase in profits is entirely due to improvements in profit productivity, and that measured business conditions actually lowered profits slightly. As a reminder, the standard profit and alternative profit optimization concepts differ only in the specification of variable output prices (standard profit concept) versus variable output quantities (alternative profit concept) in the business conditions vector X . 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), although we again caution that this breakout of productivity change into change in best practice versus change in inefficiency is the most difficult to estimate. The standard and alternative profit results qualitatively agree in all other respects. Most important and striking, both profit approaches find substantial increases in productivity during 1991-1997—annual rates of increase of 13.7% and 16.5% for standard and alternative profit, respectively—in sharp contrast to the findings of cost productivity deterioration for this subinterval.

¹⁶Of course, increases in competitive pressure may alternatively improve performance by encouraging banks to cut costs to compete, and this may have contributed to the observed productivity, best practice, inefficiency, and business condition effects, but cannot explain the observed deteriorations in productivity and best-practice performance.

5.2 Alternative explanations of the differences between the cost and profit results

We consider a number of possible explanations of the striking findings that cost productivity worsened while profit productivity improved substantially. The most likely explanation in our opinion concerns some weaknesses of the cost minimization approach as applied to the banking industry. The cost approach and other approaches that do not consider revenues may not account well for unmeasured changes in output quality or may fail to adequately capture the profit maximization goal of banks, which requires that effort be spent to raise revenues as well as reduce costs. Note that the nonparametric concepts and the government statistics suffer from these same problems. Use of the profit approaches may also help take into account unmeasured changes in the quality of banking services by including higher revenues paid for the improved quality. Over time, it seems clear that banks have offered wider varieties of financial services, such as mutual funds, derivatives, and other products of financial engineering. 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 on-line services. The cost and profit productivity results are consistent with the hypothesis that banks' goals were to maximize profits by putting effort into raising revenues as well as reducing costs, and that banks provided additional services or higher service quality that raised costs but also raised revenues by more than the cost increases.

Another possible explanation of the conflicting cost versus profit productivity results is that regulatory innovations or changes in competitive conditions not captured in our business condition variables, X , affected the liability and asset sides of banks dramatically differently. Recall that our asset approach treats all bank liabilities and equity as inputs and all assets as outputs. This potential explanation of the results seems unlikely to explain much of the findings. The main regulatory innovation of the 1990s was liberalized geographic expansion, which should affect liability and asset markets similarly, and we include a series of variables to account for this deregulation in the X vector. Also note that our measure of deposit concentration, HERF, changed little on average over time (Table 2). It seems unlikely that market power could have increased over borrowers but remained relatively constant over depositors, given that past research has found that the exercise of market power appears to occur simultaneously over both core depositors and small business borrowers in the

same local markets in which concentration is high (Berger and Hannan, 1989,1997; Hannan, 1991).¹⁷ In addition, the X_c and X_π vectors of business conditions control for market-average input prices faced, local market banking concentration, and state geographic restrictions on bank competition, and the X_π vector controls for market-average output prices faced in the standard profit concept, although the 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.

We next consider the possibility that our findings may be driven by the way we have divided up the 1984-1997 period into two intervals based on the performance patterns shown above in Table 1. The bottom panel of Table 3 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.¹⁸ It is also notable from Table 1 that most of the cost and profit performance improvements of the 1990s took place between 1991 and 1992, after which performance remained strong but improvements were much smaller. Thus, this alternative breakout of the time period may give a rigorous check of whether our findings of worsening cost productivity and improving profit productivity are robust.

The results shown in the bottom panel of Table 3 suggest that most of the main results in the top panel are robust. Again, cost productivity worsens in the early and late time subintervals and again profit productivity improves over time, with greater cost worsening and profit improvement 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. As can be seen from Table 1, there was a very sharp improvement in performance, particularly cost performance, over this short interval, most of it coming in the single year from 1991 to 1992. It is also notable that these unusual findings are reversed when the 1992-1997 subinterval is included along with the 1989-1992 subinterval—the worsening

¹⁷Some recent research also suggests that the banks with persistently high profits in the 1990s are not consistently those with high measured market power (Berger, Bonime, Covitz, and Hancock, 1999). See Berger, Demsetz, and Strahan (1999) for a more comprehensive survey of recent findings on the exercise of market power in banking.

¹⁸Note that the 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.

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; this phenomenon appears to be short-lived.

Another potential candidate for explaining our results is the inclusion of both small and large banks in the same analysis. It is often argued that small and large banks provide different products using different technologies, although balance sheet entries of small and large banks appear identical except for magnitude. For example, a \$1 billion loan issued by a large bank may be a different product requiring different monitoring and screening policies and procedures 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. Although the Fourier-flexible functional form specified for the cost and profit functions provides considerable flexibility, it cannot fully take into account differences in the products and technologies of small and large banks. To some extent, the observed effects thus far could reflect the shifting of banks into larger size classes over time, which may have different characteristics.

To investigate the possibility of important differences between small and large banks, we repeated the analysis for banks in the bottom quarter and top quarter in terms of GTA in each year. Table 4 shows these results, replicating the panels of Table 3 for the small and large banks.¹⁹ As indicated earlier, banks in the smallest quarter are unambiguously small, but the largest quarter includes medium-sized banks because there are not enough large banks alone to allow reasonably accurate estimation of our cost and profit functions. Before viewing Table 4, we caution that these results should be considered less reliable than those in Table 3. The small-bank and large-bank subsamples on which the Δ TOTAL, Δ PROD, and Δ BUSCOND effects are based each use one-quarter of the data for each year, and the small-bank and large-bank thick-frontier subsamples on which the Δ BESTPR and Δ INEFF effects are based each use only one-sixteenth of the data for each year.

The results in Table 4 suggest that our main results and conclusions about cost versus profit productivity are supported, but they also show some interesting heterogeneity in the changes in productivity and performance

¹⁹Note that we evaluate the cost and profit functions of the small-bank and of the large-bank subsamples at the average level of business conditions specific to each subsample.

between small and large banks. 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 and for both the standard and alternative approaches. The Δ BESTPR and Δ INEFF 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. 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. It is also notable that the results for the large-bank subsample were generally closer to the full-sample results shown earlier, particularly for the profit results. This is consistent with the possibility that one of the mechanisms through which the main productivity consequences may have taken place is bank mergers that shifted banks into larger size classes.

To investigate further whether merger activity may be responsible for our findings, we reestimated our cost and profit functions for each year for a **merging banks** subsample and a **nonmerging banks** subsample, and the results are shown in Table 5. 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 in any of the three prior years, and the nonmerging bank subsample includes all other banks that were in the total population used in our main results.²⁰ We strongly caution that the results for the merging banks subsample should be considered much less reliable than those in Table 3. 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). On average, merging banks tend to be larger and to have higher average profits and lower average costs than do nonmerging banks.²¹

²⁰As was true for the small-bank and large-bank subsamples, we evaluate the cost and profit functions of the merging bank subsample and the nonmerging bank subsample at the average level of business conditions specific to each subsample.

²¹On 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

The results shown in Table 5 suggest that merging banks may be responsible for much of our main findings for the 1991-1997 time interval. For this interval, 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. All of these figures are larger in absolute value than the corresponding productivity changes for the industry as a whole shown in Table 3 above.²² The changes in average costs and profits for merging banks over this interval (Δ TOTAL) are also larger than the industry figures, suggesting that the findings of large productivity changes do not simply derive from measurement error in applying our model to the relatively small number of observations. 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 than those of 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 5 suggest that merging banks may have done this more than other banks, perhaps as part of their overall strategies of refocusing their institutions.

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 prior studies that found improved profit efficiency for merging banks linked to the diversification of risks (Akhavain, Berger, and Humphrey, 1997; Berger, 1998) and with prior 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 (Berger and Humphrey, 1992; Rhoades, 1993; DeYoung, 1997; Peristiani, 1997; Berger, 1998; Rhoades,

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. Return on GTA was 0.75% in 1984, 0.77% in 1991, and 1.12% in 1997 for merging banks, and 0.69% in 1984, 0.71% in 1991, and 1.14% 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.

²²In 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.

1998). Thus, the merger efficiency literature suggests an explanation for 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. Increases in market power may also help explain part of the measured profit productivity improvement from mergers, but as discussed earlier, this is unlikely to explain much of the results given the cost findings and the controls included for prices, concentration, and state restrictions included in the X vectors. Further investigation of these specific merger effects is beyond the scope of this study.

As a final check on our results, we also investigated whether our main results were primarily due to the actions of banks that remained in the industry over the long run, or to banks that entered and exited the industry. For example, it is possible that previously existing firms did not change their behavior significantly, but that firms that entered during our time interval had lower cost productivity and higher profit productivity than those that exited. 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 acquisition by merging banks. Recall from our definition above that merging banks keep their original charters and so do not enter or exit in the year of the merger, but the banks whose charters they absorbed are counted as exiting. We examined this issue by rerunning our estimations using a balanced panel of banks that remained in the sample from 1984 through 1997. There were 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 were very similar to those reported in Table 3 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 between banks that entered and those that exited over the time interval.

6. Conclusions

This paper analyzes the sources of the recent dramatic changes in the performance of U.S. banks. We investigate three sources of these changes in performance—changes in managerial best practice, changes in cross-section inefficiency, 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 the study of firm performance because it more completely describes the economic goals of firms and their owners, who take revenues into account as well as costs. Prior research on productivity change in banking as well as government productivity statistics have almost always neglected revenues, and therefore may have yielded misleading findings. We include cost minimization and two types of profit maximization to ensure a comprehensive analysis.

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 in our opinion is weakness in the cost minimization approach, which does not account for significant revenue changes. The data are consistent with the hypothesis that banks tried to maximize profits by raising revenues as well as reducing costs. Over time, banks have offered wider varieties of financial services, such as mutual funds, derivatives, and other products of financial engineering, and provided additional convenience through more extensive branching and ATM networks, expanded availability of debit and credit cards, and a proliferation of on-line services. 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.

Further analysis also suggested 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, although we strongly caution that these results should be considered less reliable than our main results because of smaller numbers of observations. Nonetheless, these findings are consistent with the possibility that any additional services or higher service quality that raised both costs and revenues as described above 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 higher risk-higher expected return investments 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, *Review of Industrial Organization* 12, 95-139.
- Alam, I. M. S. (1998). A non-parametric approach for assessing productivity dynamics of large banks, Tulane University mimeo.
- Allen, L., Saunders, A., and Udell, G. F. (1991). The pricing of retail deposits: concentration and information, *Journal of Financial Intermediation* 1, 335-361.
- 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.
- Berg, S. A., Forsund, F., and Jansen, E. (1992). Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980-89, *Scandinavian Journal of Economics* 94, 211-228.
- Berger, A. N. (1993). "Distribution-free" estimates of efficiency in the U.S. banking industry and tests of the standard distributional assumptions, *Journal of Productivity Analysis* 4, 261-292.
- Berger, A. N. (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. (1999). "Why are bank profits persistent? The roles of regulation, market structure, and technology," Board of Governors of the Federal Reserve System Working Paper.
- Berger, A. N., Cummins, J. D., and Weiss, M. (1997). The coexistence of multiple distribution systems for financial services: The case of property-liability insurance, *Journal of Business* 70, 515-546.
- Berger, A. N., Demsetz, R. S., and Strahan, P. E. (1999). The consolidation of the financial services industry: Causes, consequences, and implications for the future, *Journal of Banking and Finance* 23.
- Berger, A. N., and DeYoung, R. (1997). Problem loans and cost efficiency in commercial banks, *Journal of Banking and Finance* 21, 849-870.
- Berger, A. N., and Hannan, T. H. (1989). The price-concentration relationship in banking, *Review of Economics and Statistics* 71, 291-299.
- Berger, A. N., Hannan, T. H. (1997). Using measures of firm efficiency to distinguish among alternative

- explanations of the structure-performance relationship, *Managerial Finance* 23, 6-31.
- 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, *Review of Economics and Statistics* 80, 454-65.
- Berger, A. N., and Humphrey, D. B. (1991). ‘The dominance of inefficiencies over scale and product mix economies in banking,’ *Journal of Monetary Economics* 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 Humphrey, D. B. (1997). Efficiency of financial institutions: International survey and directions for future research, *European Journal of Operational Research* 98, 175-212.
- Berger, A. N., Kashyap, A. K., and Scalise, J. M. (1995). The transformation of the U.S. banking industry: What a long, strange trip it’s been, *Brookings Papers on Economic Activity* 2, 55-218.
- Berger, A. N., Leusner, J. H., and Mingo, J. (1997). The efficiency of bank branches, *Journal of Monetary Economics* 40, 141-162.
- Berger, A. N., and Mester, L. J. (1997). Inside the black box: What explains differences in the efficiencies of financial institutions? *Journal of Banking and Finance* 21, 895-947.
- 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. (1996). “Productivity growth, market structure, and technological change: Evidence from the rural banking sector,” Southeast Missouri University mimeo.
- DeYoung, R. (1997). Bank mergers, X-efficiency, and the market for corporate control, *Managerial Finance* 23, 32-47.
- Elyasiani, E., and Mehdiian, S.M. (1995). The comparative efficiency performance of small and large U.S. commercial banks in the pre- and post-deregulation eras, *Applied Economics* 27, 1069-1079.
- Fixler, D., and Hancock, D. (1997). “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.
- Fixler, D., and Zieschang, K. (1998). "Economic statistics and the transmission of monetary policy to the real economy," U.S. Bureau of Labor Statistics mimeo.
- Griliches, Z. (1992). Introduction, in "Measurement Issues In the Service Sectors" (Z. Griliches, Ed.), pp. 1-22. National Bureau of Economic Research, University of Chicago Press, Chicago.
- Hannan, T. H. (1991). Bank commercial loan markets and the role of market structure: Evidence from surveys of commercial lending, *Journal of Banking and Finance* 15, 133-149.
- Hassan, I., Hunter, W. C., and Lozano-Vivas, A. (1998). Deregulation, intermediation and efficiency: Spanish experience, Federal Reserve Bank of Chicago mimeo.
- Hughes, J. P., Lang, W., Mester, L. J., and Moon, C.-G. (1996). Efficient banking under interstate branching, *Journal of Money, Credit, and Banking* 28, 1043-1071.
- Hughes, J. P., Lang, W., Mester, L. J., Moon, C.-G. (1997). "Recovering risky technologies using the almost ideal demand system: An application to U.S. banks," Federal Reserve Bank of Philadelphia Working Paper No. 97-8.
- Hughes, J. P., Lang, W., Mester, L. J., and Moon, C.-G. (1999). The dollars and sense of bank consolidation, *Journal of Banking and Finance* 23.
- 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, *Journal of Productivity Analysis* 4, 293-315.
- Hughes, J. P., and Mester, L. J. (1998). Bank capitalization and cost: Evidence of scale economies in risk management and signaling, *The Review of Economics and Statistics* 80, 314-325.
- Humphrey, D. B. (1993). Cost and technical change: Effects from bank deregulation, *Journal of Productivity Analysis* 4, 9-34.
- Humphrey, D. B., and Pulley, L. B. (1997). Banks' responses to deregulation: Profits, technology, and efficiency, *Journal of Money, Credit, and Banking* 29, 73-93.
- Hunter, W. C., and Timme, S. G. (1991), Technological change and production economies in large US commercial banking, *Journal of Business* 64, 339-362.
- 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.

- McAllister, P. H., and McManus, D. (1993). Resolving the scale efficiency puzzle in banking, *Journal of Banking and Finance* 17, 389-405.
- Mester, L. J. (1992). Traditional and nontraditional banking: An information-theoretic approach, *Journal of Banking and Finance* 16, 545-566.
- Mester, L. J., Nakamura, L. I., and Renault, M. (1998). "Checking accounts and bank monitoring," Federal Reserve Bank of Philadelphia Working Paper No. 98-25.
- Mitchell, K., and Onvural, N. M. (1996). Economies of scale and scope at large commercial banks: Evidence from the Fourier flexible functional form, *Journal of Money, Credit, and Banking* 28, 178-199.
- 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.
- Nakamura, L. I. (1993). Commercial bank information: Implications for the structure of banking, in "Structural Change in Banking" (M. Klausner, and L. J. White, Eds.), pp. 131-160. Irwin, Homewood, IL.
- Peristiani, S. (1997). Do mergers improve the X-efficiency and scale efficiency of U.S. banks? Evidence from the 1980s, *Journal of Money, Credit, and Banking* 29, 326-337.
- Rhoades, S. A. (1993). The efficiency effects of horizontal bank mergers, *Journal of Banking and Finance* 17, 411-422.
- Rhoades, S. A. (1998). The efficiency effects of bank mergers: An overview of case studies of nine mergers, *Journal of Banking and Finance* 22, 273-291.
- Sealey, C., and Lindley, J. (1977). Inputs, outputs, and a theory of production and cost at depository financial institutions, *Journal of Finance* 32, 1251-1266.
- Triplett, J. E. (1998). "Concepts of banking output: What do banks do?" Brookings Institution mimeo.
- Wheelock, D. C., and Wilson, P. W. (1996). "Technical progress, inefficiency, and productivity change in U.S. banking, 1984-1993," Federal Reserve Bank of St. Louis mimeo.
- Yuskavage, R. E. (1996). Improved estimates of gross product by industry, 1959-94, *Survey of Current Business* 76, 133-155.

TABLE 1

**Raw Data Measures of Profits and Cost
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
1984	14,095	0.0649	0.00688	1.248	0.0959	1.163	0.0528
1985	14,027	0.0430	0.00570	1.195	0.0904	1.175	0.0585
1986	13,734	0.0111	0.00424	1.153	0.0827	1.166	0.0556
1987	13,237	0.0198	0.00480	1.085	0.0784	1.178	0.0478
1988	12,639	0.0507	0.00649	1.083	0.0794	1.182	0.0420
1989	12,295	0.0705	0.00756	1.131	0.0851	1.181	0.0413
1990	11,968	0.0568	0.00656	1.130	0.0844	1.172	0.0425
1991	11,623	0.0714	0.00715	1.037	0.0802	1.183	0.0413
1992	11,238	0.1049	0.00993	0.832	0.0693	1.257	0.0337
1993	10,785	0.1160	0.01101	0.698	0.0608	1.290	0.0286
1994	10,289	0.1142	0.01072	0.699	0.0606	1.298	0.0253
1995	9,747	0.1135	0.01096	0.700	0.0646	1.285	0.0266
1996	9,303	0.1151	0.01107	0.696	0.0644	1.296	0.0277
1997	8,855	0.1148	0.01134	0.685	0.0659	1.294	0.0259

Source: Reports of Condition and Income (Call Reports).

Notes: GTA refers to gross total assets, which equals total assets plus loan and lease loss reserves and allocated transfer risk reserve (a reserve for certain foreign loans).

Nonperforming loans are loans that are past due at least 90 days or are on nonaccrual basis.

TABLE 2

**Variables Employed 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,
Prices of financial assets and liabilities are measured as interest rates.)

Symbol	Definition	1984	1991	1997
		14,095 obs.	11,623 obs.	8,855 obs.
		Mean (Std. Dev.)	Mean (Std. Dev.)	Mean (Std. Dev.)
Dependent Variables				
C	Variable operating plus interest costs, includes costs of purchased funds, deposits, and labor.	19,191 (244,441)	18,786 (217,483)	22,030 (280,864)
π	Variable profits, includes revenues from loans and securities less variable costs.	3,606 (33,750)	4,852 (32,776)	9,662 (99,310)
Variable Input Prices				
w₁	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)
w₂	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)
w₃	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)				
y₁	Consumer loans (loans to individuals).	24,885 (189,396)	33,368 (264,761)	52,145 (498,282)
y₂	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)
y₃	Real estate loans.	35,982 (321,058)	74,238 (594,825)	120,986 (1,105,862)
y₄	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 2, CON'T.

Symbol	Definition	1984	1991	1997
		14,095 obs.	11,623 obs.	8,855 obs.
		Mean (Std. Dev.)	Mean (Std. Dev.)	Mean (Std. Dev.)
Variable Output Prices (Standard Profit Function Only)				
p₁	Market-average price of consumer loans.	0.110 (0.0306)	0.104 (0.0242)	0.0963 (0.0274)
p₂	Market-average price of business loans.	0.147 (0.0428)	0.115 (0.0373)	0.0974 (0.0282)
p₃	Market-average price of real estate loans.	0.0990 (0.0172)	0.0858 (0.0142)	0.0793 (0.00967)
p₄	Market-average price of securities.	0.0755 (0.0112)	0.0612 (0.00945)	0.0447 (0.00763)
Fixed Netput Quantities				
z₁	Off-balance-sheet items (commitments, letters of credit, derivatives, etc.) measured using Basle Accord risk weights to be risk-equivalent to loans.	23,914 (529,934)	36,018 (787,436)	74,526 (1,601,009)
z₂	Physical capital (premises and other fixed assets).	3,692 (34,467)	4,732 (47,425)	6,785 (75,692)
z₃	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)

TABLE 2, CON'T.

Symbol	Definition	1984	1991	1997
		14,095 obs.	11,623 obs.	8,855 obs.
		Mean (Std. Dev.)	Mean (Std. Dev.)	Mean (Std. Dev.)
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)
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.	0.225 (0.167)	0.232 (0.158)	0.236 (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 adjusted for loan loss provisions. Since provisions are reported in the Call Report for total loans and not by loan category, we apportioned the loan loss provisions among our three loan categories based on charge-offs, which are reported by category. E.g., loan loss provisions for consumer loans = (charge-offs on consumer loans / total charge-offs) • total loan loss provisions.

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 $\rho^{jM} \equiv \sum_{i=1, i \neq j}^m \left(x^i / \sum_{i=1, i \neq j}^m x^i \right) \omega^i$, 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: that is, $w^j \equiv \sum_{k=1}^N \left(d^{jk} / \sum_{i=1}^N d^{ji} \right) \rho^{jk}$, 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 3
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				
	Δ TOTAL	Δ PROD	Δ BUSCOND	Δ BESTPR	Δ INEFF	Δ TOTAL	Δ PROD	Δ BUSCOND	Δ BESTPR	Δ INEFF	Δ TOTAL	Δ PROD	Δ BUSCOND	Δ BESTPR	Δ INEFF
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				
	Δ TOTAL	Δ PROD	Δ BUSCOND	Δ BESTPR	Δ INEFF	Δ TOTAL	Δ PROD	Δ BUSCOND	Δ BESTPR	Δ INEFF	Δ TOTAL	Δ PROD	Δ BUSCOND	Δ BESTPR	Δ INEFF
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 $_{C_{t+k}}$)^{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.

$$\begin{aligned} \Delta\text{TOTAL} &= \{\exp[f_{C_{t+k}}(\bar{X}_{C_{t+k}})] \cdot \exp[\ln u_{C_{t+k}}] / \{\exp[f_{C_t}(\bar{X}_{C_t})] \cdot \exp[\ln u_{C_t}]\} \text{ for cost;} \\ &= \{\exp[f_{\pi_{t+k}}(\bar{X}_{\pi_{t+k}})] \cdot \exp[\ln u_{\pi_{t+k}}] - \theta_{t+k} / \{\exp[f_{\pi_t}(\bar{X}_{\pi_t})] \cdot \exp[\ln u_{\pi_t}] - \theta_t\} \text{ for standard profit;} \\ &= \{\exp[f_{\text{alt}_{t+k}}(\bar{X}_{C_{t+k}})] \cdot \exp[\ln u_{\text{alt}_{t+k}}] - \theta_{t+k} / \{\exp[f_{\text{alt}_t}(\bar{X}_{C_t})] \cdot \exp[\ln u_{\text{alt}_t}] - \theta_t\} \text{ for alternative profit.} \end{aligned}$$

$$\Delta\text{PROD} = \Delta\text{BESTPR} \cdot \Delta\text{INEFF} \text{ for cost, standard profit, and alternative profit.}$$

$$\begin{aligned} \Delta\text{BESTPR} &= \{\exp[f_{C_{t+k}}(\bar{X}_{C_t})] / \exp[f_{C_t}(\bar{X}_{C_t})]\} \text{ for cost;} \\ &= \{\exp[f_{\pi_{t+k}}(\bar{X}_{\pi_t})] - \theta_{t+k} / \{\exp[f_{\pi_t}(\bar{X}_{\pi_t})] - \theta_t\} \text{ for standard profit;} \\ &= \{\exp[f_{\text{alt}_{t+k}}(\bar{X}_{C_t})] - \theta_{t+k} / \{\exp[f_{\text{alt}_t}(\bar{X}_{C_t})] - \theta_t\} \text{ for alternative profit.} \end{aligned}$$

$$\begin{aligned} \Delta\text{BUSCOND} &= \{\exp[f_{C_{t+k}}(\bar{X}_{C_{t+k}})] / \exp[f_{C_{t+k}}(\bar{X}_{C_t})]\} \text{ for cost;} \\ &= \{\exp[f_{\pi_{t+k}}(\bar{X}_{\pi_{t+k}})] - \theta_{t+k} / \{\exp[f_{\pi_{t+k}}(\bar{X}_{\pi_t})] - \theta_{t+k}\} \text{ for standard profit;} \\ &= \{\exp[f_{\text{alt}_{t+k}}(\bar{X}_{C_{t+k}})] - \theta_{t+k} / \{\exp[f_{\text{alt}_{t+k}}(\bar{X}_{C_t})] - \theta_{t+k}\} \text{ for alternative profit.} \end{aligned}$$

$$\begin{aligned} \Delta\text{INEFF} &= \{\exp[\ln u_{C_{t+k}}] / \exp[\ln u_{C_t}]\} \text{ for cost;} \\ &= \langle \{\exp[f_{\pi_{t+k}}(\bar{X}_{\pi_{t+k}})] \cdot \exp[\ln u_{\pi_{t+k}}] - \theta_{t+k} / \{\exp[f_{\pi_t}(\bar{X}_{\pi_t})] \cdot \exp[\ln u_{\pi_t}] - \theta_t\} / \\ &\quad \langle \{\exp[f_{\pi_{t+k}}(\bar{X}_{\pi_{t+k}})] - \theta_{t+k} / \{\exp[f_{\pi_t}(\bar{X}_{\pi_t})] - \theta_t\} \rangle \text{ for standard profit;} \\ &= \langle \{\exp[f_{\text{alt}_{t+k}}(\bar{X}_{C_{t+k}})] \cdot \exp[\ln u_{\text{alt}_{t+k}}] - \theta_{t+k} / \{\exp[f_{\text{alt}_t}(\bar{X}_{C_t})] \cdot \exp[\ln u_{\text{alt}_t}] - \theta_t\} / \\ &\quad \langle \{\exp[f_{\text{alt}_{t+k}}(\bar{X}_{C_{t+k}})] - \theta_{t+k} / \{\exp[f_{\text{alt}_t}(\bar{X}_{C_t})] - \theta_t\} \rangle \text{ for alternative profit.} \end{aligned}$$

TABLE 4
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 1991 = \$1 million to \$28 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 \$168 billion in assets
 Size range in 1991 = \$113 million to \$176 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 = 3,023, 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., $(\Delta\text{TOTAL}_{t,t+k})^{1/k}$].

See notes to Table 3 for definitions of the change variables.

TABLE 5
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				
	Δ TOTAL	Δ PROD	Δ BUSCOND	Δ BESTPR	Δ INEFF	Δ TOTAL	Δ PROD	Δ BUSCOND	Δ BESTPR	Δ INEFF	Δ TOTAL	Δ PROD	Δ BUSCOND	Δ BESTPR	Δ INEFF
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				
	Δ TOTAL	Δ PROD	Δ BUSCOND	Δ BESTPR	Δ INEFF	Δ TOTAL	Δ PROD	Δ BUSCOND	Δ BESTPR	Δ INEFF	Δ TOTAL	Δ PROD	Δ BUSCOND	Δ BESTPR	Δ INEFF
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				
	Δ TOTAL	Δ PROD	Δ BUSCOND	Δ BESTPR	Δ INEFF	Δ TOTAL	Δ PROD	Δ BUSCOND	Δ BESTPR	Δ INEFF	Δ TOTAL	Δ PROD	Δ BUSCOND	Δ BESTPR	Δ INEFF
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 in any of the three prior years, and the nonmerging bank subsample includes all other banks that were in the total sample used in our main results reported in Table 3. 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 k th roots of the k -period rates of change [e.g., $(\Delta$ TOTAL_{C,t+k})^{1/k}].

See notes to Table 3 for definitions of the change variables.