Business Review

Federal Reserve Bank of Philadelphia

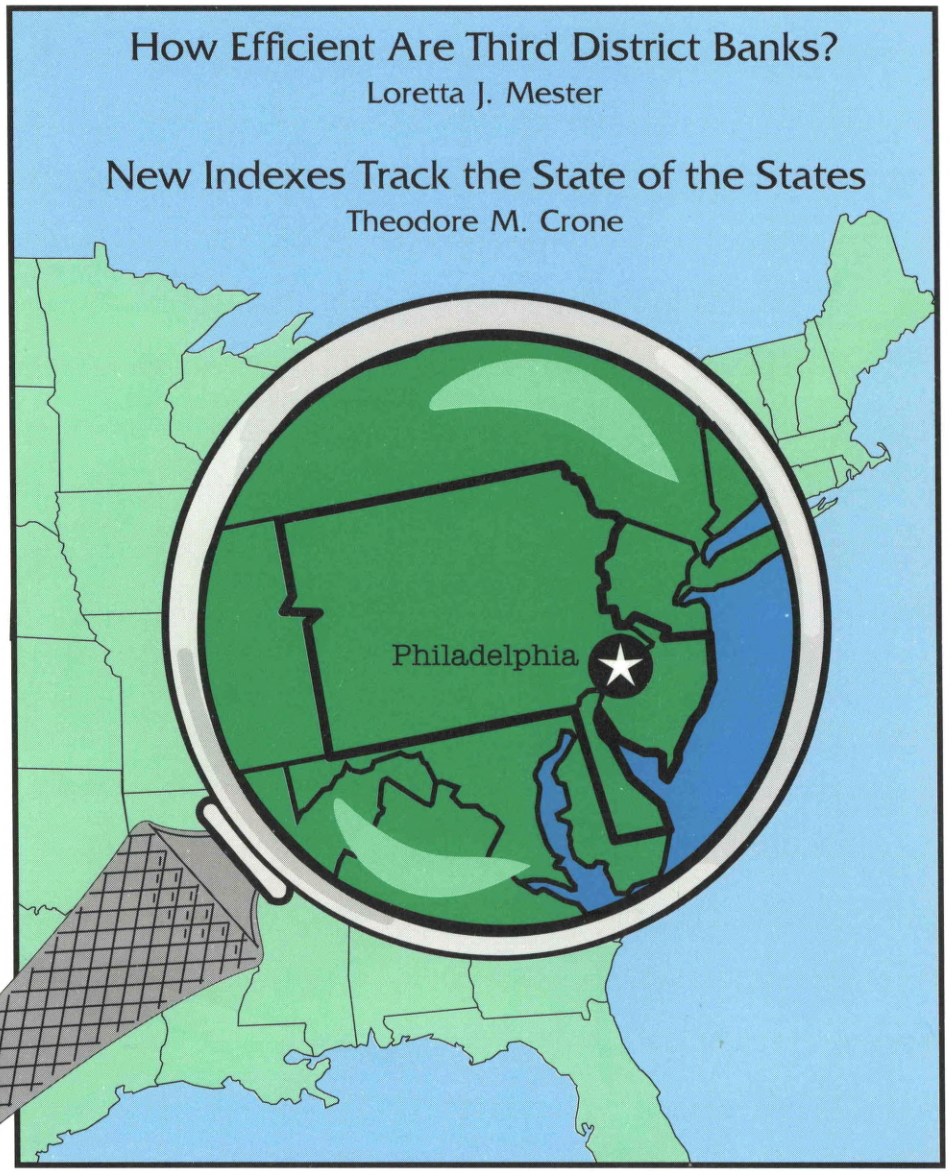
January • February 1994 ISSN 0007-7011

How Efficient Are Third District Banks?

Loretta J. Mester

New Indexes Track the State of the States

Theodore M. Crone



Business Review

The BUSINESS REVIEW is published by the Department of Research six times a year. It is edited by Sarah Burke. Artwork is designed and produced by Dianne Hallowell under the direction of Ronald B. Williams. The views expressed here are not necessarily those of this Reserve Bank or of the Federal Reserve System.

SUBSCRIPTIONS. Single-copy subscriptions for individuals are available without charge. Institutional subscribers may order up to 5 copies.

BACK ISSUES. Back issues are available free of charge, but quantities are limited: educators may order up to 50 copies by submitting requests on institutional letterhead; other orders are limited to 1 copy per request. Microform copies are available for purchase from University Microfilms, 300 N. Zeeb Road, Ann Arbor, MI 48106.

REPRODUCTION. Permission must be obtained to reprint portions of articles or whole articles. Permission to photocopy is unrestricted.

Please send subscription orders, back orders, changes of address, and requests to reprint to Publications, Federal Reserve Bank of Philadelphia, Department of Research and Statistics, Ten Independence Mall, Philadelphia, PA 19106-1574, or telephone (215) 574-6428. Please direct editorial communications to the same address, or telephone (215) 574-3805.

JANUARY/FEBRUARY 1994

How Efficient Are Third District Banks?

*Loretta J. Mester**

In recent years banks have had to operate in an increasingly competitive environment. Competitors have come from both within and outside the banking industry. Deregulation has allowed commercial banks to expand beyond their own state's borders; thus banks face competition from other commercial banks entering their market for the first time. Investment banks have also become competitors for some of the commercial bank's most creditworthy customers, who have been able to turn to the commercial paper market as a cheaper funding source than bank loans. Similarly, savers have been funneling their money into mutual funds as opposed to bank deposits in a search for a higher rate of return in the current low-deposit-rate environment. Although banks are still the main financial intermediaries in the United States, providing funding to firms and other borrowers and deposit services to savers, whether they will remain dominant in the face of increased competition depends on how efficiently they produce their outputs, that is, their loans and other financial services. Efficient banks will be able to offer more attractive loan and deposit rates to their customers and still

* Loretta J. Mester is a Research Officer in the Research Department, Federal Reserve Bank of Philadelphia, and Adjunct Assistant Professor of Finance, The Wharton School, University of Pennsylvania.

make a normal rate of return, while inefficient banks won't be able to follow suit and will, therefore, lose business. Inefficiently run banks will have to shape up, or they will be driven out of the market or acquired by other banks; indeed mergers between efficient and inefficient banks have the potential for substantial social gains via cost savings.[1] And banks that operate with lower costs can pass these savings on to their borrowers and depositors.

What can we expect for banks operating in the Third Federal Reserve District, which comprises the eastern two-thirds of Pennsylvania, the southern half of New Jersey, and Delaware? Are they operating at a high level of efficiency, or is there room for substantial improvement? Can we expect a lot of restructuring in our District as inefficient banks are driven from the market? Measuring our banks' efficiency will give us some indication of how they are likely to fare in an increasingly competitive environment.

WHAT DO WE MEAN BY EFFICIENCY?

When economists consider efficiency they typically focus on *scale* and *scope efficiency*, which concerns a bank's choice of outputs, and *X-efficiency*, which concerns a bank's use of inputs. There has been substantially more study of scale and scope efficiency in the banking and financial services industry than of X-efficiency.

Scale Efficiency. Scale efficiency refers to whether a firm is providing the most cost-efficient *level* of outputs. Let's consider a hypothetical example. Suppose firms are demanding $500 million of credit in total, and that it costs one bank $25 million to produce this volume of loans and $10 million to produce $250 million in loans. (Producing a loan involves the credit evaluation the bank must perform to determine the credit quality of the borrower along with funding the loan and monitoring the loan over its length of maturity.) Then it is more efficient to supply the $500 million of credit to the market by having two banks each produce $250 million of the loans than by having one bank produce all $500 million—the average cost of production, that is, the cost per dollar of loan, is less (4¢ versus 5¢) when each bank produces $250 million of loans than when one bank produces $500 million of loans. Society is better off with two banks producing the output rather than one bank, since the $5 million saved could be used for some other productive activity. And a bank trying to produce all $500 million of loans would find itself at a competitive disadvantage if another bank entered the market producing only $250 million of loans, because the second bank would be able to charge a lower interest rate on its loans, since its per unit production cost would be lower.

A bank is said to be producing with *constant returns to scale* if, for a given mix of products, a proportionate increase in all its outputs would increase its costs in the same proportion; this is also the point where the average cost of production is the least. A bank is operating with *scale economies* if a proportionate increase in its outputs would lead to a less than proportionate increase in cost—the bank could produce more efficiently by increasing its output level. A bank is operating with *scale diseconomies* if a proportionate decrease in its outputs would lead to a more than proportionate decrease in costs—the bank could produce more efficiently by reducing its output level.

At output levels where there are scale economies, an increase in outputs would reduce the average cost of production, since it costs proportionately less to produce at a larger scale.

[1] Although Berger and Humphrey (1992b) found little in the way of cost-efficiency benefits on average from mergers in the 1980s between banks with assets over $1 billion, as they discuss, this is likely because the aim of such mergers was to increase asset growth and geographic market extension rather than to increase cost efficiency. This seems to be changing in the 1990s, as merger participants have been setting cost-cutting goals when announcing their mergers.

One reason it might cost less per unit to produce at a larger scale is that there may be large fixed costs in the production technology that are independent of the level of output produced. For example, in banking, the cost of the computers used to keep track of accounts can be spread over a larger number of accounts as the scale of operations increases, so that the per unit cost of production falls. Another potential source of scale economies is specialization. Larger firms may permit employees to specialize in one task, and this specialization may also lead to more efficient production.[2] In most industries, including banking, firms find that at a certain output volume (for a given mix of products), average cost stops declining. For example, at a large enough volume, the fixed costs of production will become insignificant relative to the cost of producing additional units of output, so that scale economies are exhausted. And if scale is increased beyond a certain level, average costs begin to rise. Of course, depending on the production technology, there may be a broad range of output levels (for any product mix) that firms can produce at minimum average cost. In other words, banks of different asset sizes may be equally competitive with one another, since their average costs are similar.

Scope Efficiency. Scope efficiency refers to whether a firm is producing the most cost-efficient *combination* of products. Banks produce more than one product—for example, most commercial banks produce a variety of different loans, like commercial and industrial loans, commercial real estate loans, residential mortgages, student loans, etc. To the extent that different types of loans have different default rates or other characteristics and to the extent that they aren't used to fund the same activities, they constitute different outputs of the bank.[3] Thus, in addition to choosing the most cost-efficient scale of operations, the bank must also choose the combination of products it will produce. For a given level of outputs, the per unit cost of production may be smaller if the bank produces all of the products rather than specializing in just a few of them, or it might be more efficient to specialize. There are *scope economies* if the cost of producing a given level of outputs is lower when a bank produces all the products than if the products are divided up into specialized banks. There are *scope diseconomies* if the costs are lower when specialized banks produce the various outputs.

There are several potential sources of scope economies.[4] One is the sharing of inputs to produce several outputs. For example, the same group of tellers might handle both checking and savings accounts, or information on a firm produced in a credit evaluation for a mortgage can be used if the firm wants a business loan as well. Therefore, it would be cheaper for the same bank to handle both types of accounts and to extend both loans than to duplicate the tellers and credit check at another bank. Similarly, excess capacity on the bank's computer may allow it to increase the scope of products it produces as well as its scale. Thus, there is an interconnection between scale and scope economies—the fact that the bank is able to process various types of loans on its computer (many products) enables it to increase its scale and take advantage of any scale economies. Of course, there may be a point at which producing many products will increase the bank's unit costs. For example, it may take a more elaborate hierarchical management structure to produce different product lines (some-

[2]Mester (1987) discusses the sources of scale economies in more detail.

[3]Large banks also engage in many off-balance-sheet activities, like underwriting, letters of credit, and loan guarantees.

[4]Mester (1987) describes the sources of scope economies in more detail.

times this is mandated by regulation—e.g., equities underwriting and commercial lending must be done in separate subsidiaries of a bank holding company), and this hierarchical structure can increase production costs.[5]

X-Efficiency. If all firms in the industry are producing the level and combination of outputs that minimize the average cost of production, the total cost of producing the industry's output is minimized, and the industry is producing an efficient combination and level of products, *provided each firm is using its inputs efficiently. X-efficiency* refers to whether a firm is using its inputs, like labor and capital, in a cost-effective manner—that is, for a given level and mix of outputs, is a bank producing them in the cheapest way possible? If not, the bank is either wasting some of the inputs it has purchased, or it is using the wrong combination of inputs to produce its outputs. *Technical inefficiency* refers to using proportionately too much of all inputs and is just pure waste. For example, the bank may be using too many tellers and too many branches to produce its products—it might be able to scale back its inputs and produce the same amount of service. A bank that is technically inefficient is said to be operating within its "production possibility frontier." (The production possibility frontier indicates the maximum amount of output that can be produced with a given amount of inputs.) But wasting resources is not the only way to inefficiently use inputs. A bank might be able to produce a given amount of loans and other financial services by combining its inputs—including labor, physical capital, and deposits—in different proportions than it currently is doing. For example, a large bank might be able to supply its output more cheaply by substituting ATMs for tellers—while the fixed costs of setting up an ATM are high, the cost per transaction for an ATM is lower than that for a human teller—so larger banks might benefit by using more ATMs than tellers. *Allocative inefficiency* refers to using the wrong combination of inputs to produce a given output level and product mix—an allocatively inefficient bank is operating on its production possibility frontier—that is, given the inputs it has chosen, it is producing as much output as possible—but the bank could lower its costs of producing that output by selecting a different input mix. Of course, a bank can be both technically and allocatively inefficient.

A bank that is operating in an inefficient manner might be doing so because its manager isn't on top of things, but managerial inability isn't the only source of X-inefficiency. It's possible that a bank manager has goals that differ from those of the bank's shareholders. Shareholders want to maximize the stock market value of the bank, and so its long-run profits. Thus, shareholders want the bank to minimize its cost of production. But bank managers might be interested in something other than cost-minimization. For example, managers might desire a larger staff because they think it gives them more prestige within the banking community. Thus, a bank might use an inefficient combination of inputs (more labor than is necessary) to produce its services. Such "expense-preference" behavior on the part of managers has been found in studies of commercial banks and savings and loans.[6] The bank's choice of the products it wishes to produce might also be driven less by cost considerations than by managerial desires to run a particular type of bank.

Survival. One might question how inefficient banks are able to continue operating. In the usual economic models of competitive markets, competitive forces are thought to drive

[5]See Mester (1991) for further discussion of diseconomies of scope in hierarchies.

[6]Mester (1989) discusses the conflicts between owners and managers in financial firms and the empirical evidence.

such inefficient banks out of the market. More efficient banks are able to produce at a lower cost. In a competitive market, the efficient bank would share its cost savings with its customers in the form of lower interest rates on loans and/or higher deposit rates. This would attract borrowers and depositors away from inefficient banks, since the inefficient banks couldn't match the lower prices without making a loss. The inefficient banks would eventually be forced out of the market.

But banking has been a regulated industry; competitive pressures have not been as strong as they might have been. For example, prior to the 1980s, regulations restricted bank holding companies from establishing banks in more than one state, and there are still restrictions on banks establishing branches across state lines. Such restrictions reduce the number of potential competitors, making it easier for inefficient banks to survive.[7] Similarly, laws that restrict hostile takeovers make it less easy for more efficient banks to gain control of their less efficient counterparts. On the customer side, there is empirical evidence that bank customers have found it costly to switch banks (see Calem and Mester, 1993); thus, it has been difficult for efficient banks to attract customers with lower prices.

But inefficient banks will find it less easy to survive in the future as entry barriers fall. States began passing laws in the 1980s that authorize interstate banking for bank holding companies. All but two states (Hawaii and Montana) now allow bank holding companies from at least some other states to acquire in-state banks. In April 1992, the Office of Thrift Supervision adopted a rule allowing full nationwide branching for healthy federally chartered savings and loans. According to the *American Banker* (August 2, 1993) four states—New York, North Carolina, Oregon, and Alaska—have passed reciprocal interstate branching laws permitting a state-chartered commercial bank that is not a member of the Federal Reserve System to become a branch of a bank in any other state that has an identical law. Although interstate branching hasn't been authorized for national banks as yet, Congress has considered several proposals to permit it, and the topic is likely to remain on the agenda. In addition, the Federal Reserve has taken the position that it will treat hostile bids no differently from friendly bids in assessing whether to permit a takeover. Nonbank competition is also picking up. According to the flow of funds accounts, banks' share of total U.S. financial assets has shrunk to less than 25 percent from over 35 percent in 1977.[8] And foreign bank competition is heating up too; the North American Free Trade Agreement (NAFTA) should also increase competition. Increased competitive pressures will make it more difficult for inefficient banks to survive, as will anything that reduces the costs customers face in switching to low-cost banks. For example, the Truth in Savings Act, part of the Federal Deposit Insurance Corporation Improvement Act of 1991, requires banks to report the terms of their deposit accounts in all advertisements for these accounts, making it easier for customers to shop for the best rates. Inefficient banks will find it more difficult to keep well-informed customers.

MEASURING EFFICIENCY: THE METHODOLOGY

Outputs and Inputs. Studies of bank efficiency are based on an analysis of banks' cost structure, that is, the relationship between

[7]Calem (1993) discusses the benefits of allowing banks to branch across state lines.

[8]The flow of funds accounts, published by the Board of Governors of the Federal Reserve System, provide data on financial assets and liabilities outstanding by sectors of the economy and by type of transaction.

banks' costs and their output levels, given the input prices they face. Thus, the first step in measuring efficiency in banking—scale and scope efficiency and X-efficiency—is to determine a bank's outputs and inputs. There is some disagreement in the literature over what a commercial bank is actually producing. Two general approaches have been taken: the "production" approach and the "intermediation" approach (also called the "asset" approach). The production approach focuses on the bank's operating costs, that is, the costs of labor (employees) and physical capital (plant and equipment). The bank's outputs are measured by the *number* of each type of account, like commercial and industrial loans, mortgages, deposits, because it is thought that most of the operating costs are incurred by processing account documents and debiting and crediting accounts; inputs are labor and physical capital. The intermediation approach considers a financial firm's production process to be one of financial intermediation (the borrowing of funds and the subsequent lending of those funds). Thus, the focus is on total costs, including both interest and operating expenses. Outputs are measured by the dollar volume of each of the bank's different types of loans, and inputs are labor, physical capital, and deposits and other borrowed funds.[9] Luckily, the empirical results on scale and scope efficiency do not seem to be very sensitive to which approach is taken.

Theoretically, to compare one bank's efficiency to another's, we would like to compare each bank's cost of producing the *same* outputs. For banks, significant characteristics of loans are their quality, which reflects the amount of monitoring the bank does to keep the loan performing, and their riskiness. Unless these characteristics are controlled for, one might conclude a bank was producing in a very efficient manner if it were spending far less to produce a given output level, but its output might be highly risky and of a lower quality than that of another bank. It would be wrong to say a bank was efficient if it were scrimping on the credit evaluation needed to produce sound loans. Although previous efficiency studies have failed to compare the costs of producing outputs of equal quality and risk, the study of Third District banks described below does so.

Scale and Scope Efficiency Studies. Most of the studies interested in measuring scale and scope efficiency for a particular sample of banks have estimated an *average practice cost function*, which relates a bank's cost to its output levels and input prices. The technique implicitly assumes that all banks in the sample are using their inputs efficiently, that is, there is no X-inefficiency, and they are using the same production technology. Of course, it recognizes that data are typically measured with error and that there might have been unpredicted factors

[9]A slight variation on the intermediation approach, which has been used in some studies, is to distinguish between transactions deposits, which are treated as an output, since they can serve as a measure of the amount of transactions services the bank produces, and purchased or borrowed funds (like federal funds or large CDs purchased from another bank), which are treated as inputs, since the bank does not produce services in obtaining these funds. The strict intermediation approach would consider the transactions services produced by the bank as an intermediate output, something that must be produced along the way toward the bank's final output of earning assets. Hughes and Mester (1993) empirically tested whether deposits should be treated as an input or output and found that they should be treated as an input in their study.

Another approach that has been taken less often is the "value-added" approach, which considers all liabilities and assets of the bank to have at least some of the characteristics of an output. See Berger and Humphrey (1992a) for further discussion.

Still another approach, taken in Mester (1992), is to consider the bank's output to be its loan origination and loan monitoring services.

See Humphrey (1985) and Berger and Humphrey (1992a) for further discussion of the different approaches to measuring bank output.

that affected a bank's cost over the period when the data were collected, like an unusually large amount of computer down time or up time (bad and good luck) or extraordinary sick leave. Thus, no bank is expected to lie precisely on the estimated cost function; instead the function indicates what, on average, it costs a bank facing a particular set of input prices to produce a particular bundle of outputs. Some banks will produce the given output at a slightly higher cost and others at a slightly lower cost than is indicated by the estimated cost function.

Most studies have focused on smaller banks, with assets less than $1 billion. These studies, others that included banks of all sizes, and another study that included all banks with assets greater than $100 million found that the average cost curve is relatively flat, with scale economies exhausted somewhere between $75 million and $300 million in assets.[10] This is a relatively small size when you consider the size distribution of U.S. banks. While in 1992 about 90 percent of the 11,461 FDIC-insured commercial banks in the United States had less than $300 million in assets, these banks held only 20 percent of total bank assets. Fifty-one banks had assets over $10 billion, and the largest, Citibank, had over $150 billion. Thus, the studies of scale economies suggest that only small banks are operating with unexploited economies of scale and could become more efficient producers by expanding their output size. Moreover, the measured scale economies for these small banks are usually fairly small: a 1 percent increase in all output levels typically leads to about a 0.95 percent increase in total cost, which means a 0.05 percent decrease in the average cost of producing the bundle of outputs. A handful of studies have focused solely on large banks with assets over $1 billion. Some found scale economies at very large banks—the minimum of the average cost curve usually was found to lie between $2 billion and $10 billion in assets. But here again, measured economies were not very large. On the whole, these studies concluded that there wasn't much in the way of cost gains to be made by changing the scale of operations at the typical bank.[11] Similarly, although there are exceptions, most studies have found little evidence of economies or diseconomies of scope between the products banks currently produce. Hence, there is little evidence that changing the typical bank's product mix would significantly influence its cost of production.[12]

X-Efficiency Studies. More recent studies have focused on measuring not only scale and scope economies but also the degree of X-inefficiency in banking. As with scale and scope efficiency, we start with a set of banks that are using the same production technology for creating output. The technique is to estimate a *best practice cost function*—that is, the predicted cost function of banks that are X-efficient—and then measure the degree of inefficiency relative to this best practice technology. Two common methodologies are *data envelopment analysis* (DEA) and *stochastic econometric cost frontier analysis*.[13]

[10]Berger and Humphrey (1992b), Evanoff and Israilevich (1991), Clark (1988), and Mester (1987) summarize the results of the studies.

[11]This isn't to say that banks operating at a significant distance from optimal scale couldn't become more efficient by changing their operating scale. See Evanoff and Israilevich (1991) for more discussion on this point.

[12]This is not to say that deregulation that permits banks to expand the types of products they can offer (e.g., equities underwriting) could not enable banks to take advantage of potential scope economies.

[13]There are other techniques for deriving efficiency measures, including so-called "thick frontier" analysis and "shadow price" models. Evanoff and Israilevich (1991) describe these techniques.

A simpler method to compare the efficiency of banks is to use peer-group analysis. Certain cost ratios are com-

DEA uses the data on costs, outputs, and input prices for a sample of banks and determines, for each output bundle and set of input prices, the bank in the sample that spends the least to produce the output bundle at the given input prices—this is the "best practice" (that is, most efficient) bank for that output/input price combination. (If no bank in the sample produces a particular combination, then a "best practice" bank for the combination is approximated based on "best practice" banks producing similar combinations that do show up in the sample.) A bank's relative inefficiency is then measured by the ratio of its own cost compared with the cost of the "best practice" bank that faces the same input prices and produces the same output bundle. The technique is called data *envelopment* analysis because the data on best practice banks "envelop" the data from the rest of the banks in the sample. One benefit of DEA is that it doesn't posit a particular functional form for the best practice banks' cost function—it is more flexible. But a serious drawback of the technique is that it does not allow for any error in the data—banks that have been lucky or whose costs have been undermeasured will be labeled as most efficient and other banks will look relatively less efficient in comparison. Similarly any unfavorable influence beyond the bank's control will be attributed to inefficiency.

Cost frontier analysis does not have to assume data are measured without error. Instead, a bank is labeled as inefficient if: (1) its costs are higher than the costs predicted for an efficient bank producing the same outputs and facing the same input prices *and* (2) the difference cannot be explained by statistical noise, e.g., measurement error or luck.[14] To obtain the *cost frontier*, that is, the relationship between costs, outputs, and input prices for the *efficient* banks, statistical techniques (that is, regression analysis) are used to obtain the best fitting curve through the data, just as they are used to obtain the average practice cost function usually employed in the scale and scope economies studies. The difference is that the cost frontier indicates what, on average, it costs an efficient bank facing a particular set of input prices to produce a particular bundle of outputs, while the average practice function applies to all banks. A particular bank's cost will deviate from that predicted by the cost frontier for two reasons: first, there will be statistical noise, or unpredicted factors, that affected the bank's costs—either positively or negatively—compared with an efficient bank's costs; second, the bank may not be X-efficient—hence its costs will be higher than those of efficient banks. The statistical technique used to obtain the cost frontier also provides information on these two types of deviations in the sample. The second deviation is always positive, since inefficient banks' costs are always higher than efficient banks' costs. This "one-sided" deviation can be used to obtain measures of any particular bank's inefficiency or the average level of inefficiency in the sample of banks. (As with the average practice cost function, no efficient bank is expected to lie precisely on the estimated cost frontier. Hence, the point estimate of inefficiency for these banks will be small but not

pared for banks that are considered to be similar in the types of customers they serve and products they produce. These ratios might include operating expenses per dollar volume of assets, number of employees per dollar volume of loans, or expenses attributed to commercial loans per volume of commercial loans. The Functional Cost Analysis (FCA) data collected by the Federal Reserve System permit such an analysis. The drawback of the cost ratio approach is that it cannot control for differences in banks' product mix or in the input prices banks face, which influence bank costs, and it cannot give an overall measure of efficiency. Also, the FCA program is voluntary and the sample is skewed toward smaller banks. And a bank's allocation of cost into various lines of business may require some arbitrary division of fixed or shared costs.

[14] Again, the banks in the sample are assumed to be using the same production technology in producing their outputs.

zero.) Once the cost frontier is estimated, one can also estimate scale and scope economies for banks operating efficiently.[15]

One drawback of cost frontier analysis compared with DEA is that it does require the researcher to make more assumptions about the form of the frontier and the errors; hence, it is less flexible. However, this is a less serious problem than DEA's inability to allow for any noise in the data.[16] Therefore, I use frontier analysis to analyze efficiency of banks in the Third Federal Reserve District. Another potential problem with frontier analysis is that if the researcher misspecifies the cost function to be estimated or omits factors that affect cost, this may be attributed incorrectly to inefficiency. Current research is expanding on the methodology by trying to actually model the inefficiency rather than rely on deviations from the frontier to capture inefficiency. This has great potential, since it would more readily indicate the causes of inefficiency. (See Faulhaber, 1993.)

The handful of frontier studies (including stochastic econometric and thick frontier methodologies), in general, used data from the 1970s and 1980s, and have found X-inefficiency on the average of about 20 to 30 percent in banking (see Evanoff and Israilevich, 1991). That is, elimination of X-inefficiency at the average bank could produce about a 20 to 30 percent cost savings, making this a much more serious source of inefficiency than scale and scope inefficiency. Not surprisingly, since DEA attributes any statistical noise to inefficiency, the estimates of inefficiency are higher from these studies—on the order of 20 to 50 percent. These results suggest that there is substantial room for improvement at the average bank in the United States, and the average bank will have to cut costs considerably or will have to leave the industry via merger or failure as competitive pressures increase. Is the same true of Third District banks?

EFFICIENCY OF THIRD DISTRICT BANKS

I used the cost frontier approach to study the efficiency in 1992 of commercial banks operating in the Third Federal Reserve District, which comprises the eastern two-thirds of Pennsylvania, the southern half of New Jersey, and the entire state of Delaware. Since I wanted to estimate the cost frontier of standard commercial banks that are using the same production technology, some banks were omitted from the sample.[17] The sample of 214 banks included all the Third District banks except the special purpose banks in Delaware (legislated under the Financial Center Development Act and the Consumer Credit Bank Act—thus, we excluded Delaware's credit card banks), de novo banks (that is, banks less than five years old, which have start-up costs that more mature banks do not have), and three very large banks (which very likely use different production techniques than the other banks).[18] The median asset size

[15]A more technical explanation of the frontier methodology is contained in Mester (1994).

[16]Moreover, there are ways of relaxing some of the maintained assumptions of stochastic frontier analysis and achieving more flexibility, depending on the available data. For example, using panel data—that is, data from several periods (years, quarters, etc.) on the same sample of banks—allows some of the assumptions regarding the error structure to be relaxed. See Schmidt and Sickles (1984).

[17]Since efficiency is measured relative to the cost frontier, it is important that all banks in the sample have access to the same frontier; hence they should be using the same technology. (Whether one technology is better than another is a separate issue.) One advantage to restricting the sample to the Third District rather than using a U.S. sample is that banks in the Third District are likely to have more in common with each other, thus making it more likely they are using the same production technology. It should be remembered that the results presented below apply only to the 1992 period. Since branching restrictions have only recently been eliminated in Pennsylvania—branching throughout the state became totally unrestricted only on March 4, 1990—more years of data were not included in the study.

of banks included was $144 million, and the average asset size was $325 million.

The intermediation approach was used to determine bank outputs and inputs. Three outputs were included: real estate loans, commercial and industrial plus other loans, and loans to individuals. Each of these was measured by the average dollar volume that the bank held in 1992. These three outputs account for just about all of a bank's nonsecurities earning assets. The average volume of each of these three outputs at banks in the sample was about $120 million, $52 million, and $31 million, respectively. Thus, about 60 percent of the average bank's loan portfolio is in real estate, about 25 percent is business loans, and the rest is loans to individuals.

The inputs (whose prices are used to estimate the cost frontier) are labor, physical capital, and borrowed money (including deposits, federal funds, and other borrowed money) used to fund the outputs. To account for the quality of the banks' outputs and bank risk (and so to avoid labeling as efficient banks that are not monitoring their loans), a bank's volume of nonperforming loans and the volume of its financial capital are included as arguments in the cost function.[19] The volume of nonperforming loans relative to the level of bank output is inversely related to quality: the higher the bank's nonperforming loans for a given volume of loans, the less resources the bank likely spent on monitoring its loan portfolio.[20] The higher the bank's level of financial capital relative to the level of output, the lower the bank's probability of failure and so the bank's interest costs. Financial capital is also included because capital can be used as a funding source for loans.

Scale and Scope Economies at Efficient Third District Banks. The estimated average cost frontier for Third District banks seems to be quite flat. The efficient bank producing the average level of each output and facing the average input prices is producing with constant returns to scale. That is, a 1 percent increase in the level of all outputs would lead to about a 1 percent increase in costs. (See the Table. The first line of the table's top panel, *Average Inefficiency Measures,* shows the average bank's point estimate of the scale economies measure, indicating the percentage increase in cost from a 1 percent increase in all outputs, holding quality and risk constant; it is statistically insignificant from one.) Moreover, over the entire size range of banks operating in the District, efficient banks are operating with constant returns to scale. The first line of the table's middle panel, *Scale and Scope Economies over Different Sized Banks,* shows the scale economies measures for the average efficient bank in each of four size categories. (Although the point estimates suggest decreasing average costs, the scale economies measures are sufficiently close to one that a flat average cost curve cannot be ruled out statistically.) Therefore, there do not seem to be many cost efficiency

[18]If the banks in the District are ordered by asset size, the sizes grow relatively smoothly from about $13 million to about $3.8 billion; then there is a jump to $7.8 billion, then to $9.3 billion, and then to $16 billion. Since there is empirical evidence that very large banks use a different production technology than other banks (e.g., findings of scale economies differ for small and large banks), and large banks also produce different outputs from small banks (e.g., they have more off-balance-sheet business), these three largest banks were not included in the sample.

[19]The translog functional form was assumed for the cost function; the two-sided error representing statistical noise was assumed to have a normal distribution; the one-sided error representing X-inefficiency was assumed to have a half-normal distribution. Interested readers may consult Mester (1994) for further details on the study's setup.

[20]Nonperforming loans are loans that are 30 or more days past due but still accruing interest plus loans that are not accruing interest. While the macroeconomy can affect nonperforming loans, the effect is felt equally across banks. It is the differences in nonperforming loans across banks that capture differences in quality across banks.

TABLE

Average Inefficiency Measures

Scale Economies[a]	0.95%
Scope Economies[b]	0.37%
X-Inefficiency[c]	7.90%

Scale and Scope Economies over Different Sized Banks

	Banks with Assets Under $72 Million (53 banks)	Banks with Assets Between $72 Million and $144 Million (54 banks)	Banks with Assets Between $144 Million and $280 Million (53 banks)	Banks with Assets Over $280 million (54 banks)
Scale Economies[a]	0.89%	0.92%	0.94%	0.99%
Scope Economies[b]	0.006%	0.22%	0.50%	1.10%

Bank-Specific X-Inefficiency Measures[d]

	Range of X-Inefficiency over All Banks in Each Subsample	Average X-Inefficiency over All Banks in Each Subsample[e]
Pennsylvania (182 banks)	2.94% to 19.15%	7.74%
New Jersey (26 banks)	3.71% to 22.97%	9.34%
Delaware (6 banks)	3.69% to 8.58%	6.32%

[a]The scale economies measure is $(\partial ln\ C/\partial ln\ y_1)+(\partial ln\ C/\partial ln\ y_2)+(\partial ln\ C/\partial ln\ y_3)+(\partial ln\ C/\partial ln\ k)+(\partial ln\ C/\partial ln\ q)$ where C is the predicted cost of producing the average output bundle (in the specified bank size category) at the average input prices, y_i is the volume of output i, k is the level of financial capital, and q is the volume of nonperforming loans. The measure indicates the percentage increase in costs from a 1 percent increase in each output level, holding risk and quality constant. Constant returns to scale is indicated if the measure is insignificantly different from one; decreasing returns to scale is indicated if the measure is significantly greater than one; increasing returns to scale is indicated if the measure is significantly less than one.

None of the scale economies measures is significantly different from one, so there is no evidence of scale economies or diseconomies; that is, there are constant returns to scale.

[b]The scope economies measure is $\{[C(y_1,y_2^m,y_3^m)+C(y_1^m,y_2,y_3^m)+C(y_1^m,y_2^m,y_3)]-C(y_1,y_2,y_3)\}/C(y_1,y_2,y_3)$ where y_i is the volume of output i, y_i^m is the least amount of output i produced by any bank in the sample, and $C(\bullet)$ is the predicted cost of producing an output bundle at the average input prices. The scope measure gives the percentage increase in cost of dividing the bank's products among three banks, each of which is relatively specialized in one of the three outputs. A statistically positive scope measure indicates there are economies of scope between the three outputs; a statistically negative scope measure indicates there are diseconomies of scope between the three products.

None of the scope measures is significantly different from zero, so there is no evidence of scope economies or diseconomies.

[c]The X-inefficiency measure is significantly different from zero (at the 10 percent level). This measure is $E(u_i)$ where u_i is the one-sided component of the composed error term in the frontier regression. See Mester (forthcoming).

[d]The bank-specific inefficiency measure is $E(u_i \mid \varepsilon_i)$ where u_i is the positive component of the composed error term ε_i of the frontier regression. See Mester (forthcoming).

[e]Regression results indicate that while the average point estimates differ across states, once bank characteristics are controlled for there is no statistical difference in inefficiency across states.

gains to be made from Third District banks' changing their sizes, and these results are much like those obtained in studies using U.S. samples.[21]

The scope economies statistics give the percentage increase in cost if the bank's three outputs were divided up and produced in three banks, each of which is relatively specialized in one of the outputs.[22] These measures indicate that there is no evidence of economies or diseconomies of scope at the average efficient bank in the sample nor at banks in different size categories, since the measures are statistically insignificant from zero. (See the Table.) Thus, there do not appear to be many cost efficiency gains to be made by a bank's changing its loan mix (which for the typical bank in the sample is weighted toward real estate loans).

X-Inefficiency at Third District Banks. The cost frontier technique allows one to estimate the average level of X-inefficiency for the entire sample of banks and also bank-specific levels of inefficiency. The bank-specific measures can then be averaged by state to indicate the average level of inefficiency of banks in each of the three states in the District. As shown in the table's top panel, *Average Inefficiency Measures,* and in the bottom panel, *Bank-Specific X-Inefficiency Measures,* X-inefficiency at banks in the Third District runs in the 6 to 9 percent range. In other words, given its particular output level and output mix, if the average bank were to use its inputs as efficiently as possible, it could reduce its production cost by roughly 6 to 9 percent. The average annual cost of output production at banks in the sample was about $12 million, so a 6 percent reduction in cost could potentially add about $720,000 to bank profits, which, given the average bank's size of $325 million in assets, constitutes a potential increase of 0.2 percent in before-tax return on assets, or about 0.15 percent in after-tax ROA. This isn't a trivial amount, as the average bank in the District had an after-tax ROA of 1 percent in 1992. In competitive markets not all of this gain would be retained by the bank—the savings would be passed on to customers in the form of lower loan rates and higher deposit rates. Regardless of who receives the savings—banks or their customers—society gains, since the savings created by increased efficiency can be used for other productive purposes.

Of course, not all banks are the "average" bank. The figure, *Third District Inefficiency Distribution,* indicates the number of banks in the sample that fall into different inefficiency ranges. As you can see, while the distribution is weighted in the 6 to 9 percent range, some banks are quite efficient but others show a good deal of inefficiency (as high as 23 percent). When compared with results of other studies using U.S. samples that found average X-inefficiency on the order of 20 to 30 percent, Third District banks seem to be performing better. It is difficult to determine whether this is a statistically significant difference, however. It might just reflect that the Third District study is based on more recent data, or it might be because banks in the U.S. samples are more diverse, making efficiency measurement more difficult.[23] In any case, as with U.S. banks in general, it appears that many Third District banks have room for improvement.

Characteristics of Inefficient Banks. Ultimately, we'd like to be able to say what banks can do to increase their efficiency. For each bank in the sample, the cost frontier analysis

[21]The average scale measure for the sample indicates that a 1 percent increase in output would yield a 0.95 percent increase in cost, which translates into a trivial potential increase in the average bank's return on assets.

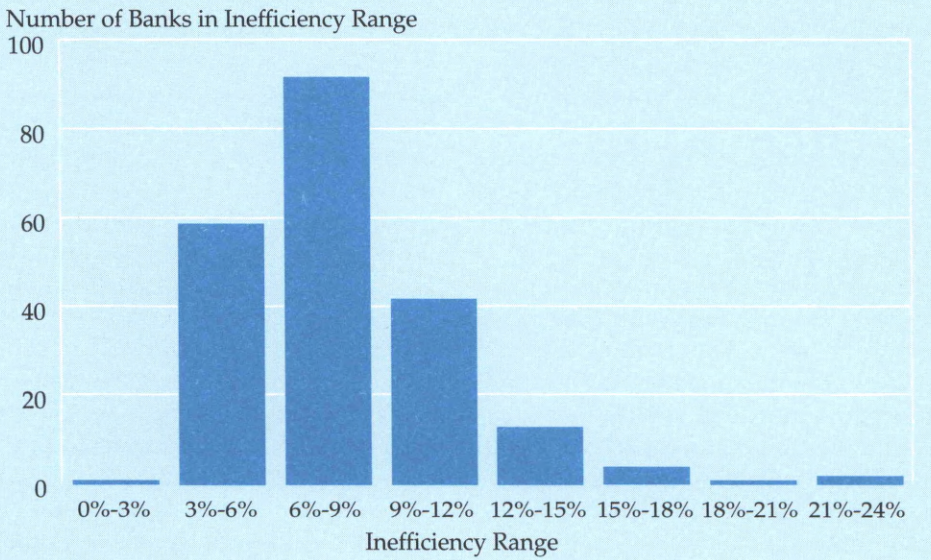
[22]This is the "within-sample degree of scope economies" as defined in Mester (1991).

[23]It might also be because Third District banks use a different production technology than other U.S. banks.

FIGURE

Third District Inefficiency Distribution

(214 Banks in the Sample)



provides a point estimate of its level of X-inefficiency. Perhaps the best way to determine what banks should do to raise efficiency is to go on site to the banks that are identified as most efficient in the study and see what they are doing differently from the banks that are least efficient. A simpler first step is to see if there are any aspects of the banks that seem to be related to their degree of inefficiency. (Of course, a relationship need not imply causality. That is, we are not saying these characteristics cause inefficiency, only that they seem to be more prevalent in inefficient banks.[24]) Simple correlations between the inefficiency measures and characteristics of the banks can be calculated, and the inefficiency measures can be regressed on bank characteristics to get an idea of how the inefficient and efficient banks in the sample differ.[25]

[24]Another reason to interpret the results as providing information on correlation only instead of causality is that there may be some endogeneity, since the characteristics are for the same period as the inefficiency measures. Causality may run from inefficiency to the characteristics instead of the other way around. For example, inefficient firms may choose to invest in real estate rather than investing in real estate leading to inefficiency.

[25]The regression involved estimating a logistic equation relating the bank-specific inefficiency measure to the following regressors: charter type (federal vs. state), holding company status (member of a holding company or not), member of the Federal Reserve System or not, number of branches, total assets, location in Pennsylvania, location in New Jersey, location in Delaware, total qualifying capital/assets, return on assets, volume of uninsured deposits/total deposits, construction and land development loans/total loans, real estate loans/total loans, loans to individuals/total loans, and year opened. See Mester (1994) for further details.

The simple correlation, which does not hold constant the other characteristics, and the regression results, which do hold constant other characteristics of the bank, indicate that inefficient banks in the District tend to be younger than more efficient banks. This might be evidence that banking involves "learning by doing," or it might indicate that more efficient banks are more likely to survive. (Recall that the de novo banks were not included in the sample, so the result probably doesn't merely reflect younger banks' higher start-up costs, for example, the costs of establishing customer relationships.)

Even though the point estimates show differences in inefficiency among banks in the three states, once other bank characteristics are controlled for, there is no statistically significant difference in inefficiency across the states.[26] Similarly, there is no evidence that larger banks are more or less X-efficient than smaller banks. This result, coupled with our results on scale economies, suggests that banks of all sizes in our District can be equally competitive when it comes to cost efficiency.

Among the statistically significant relationships, one of the more interesting is the negative relationship between inefficiency and the capital-asset ratio.[27] This result should not be interpreted as saying that if a bank increases its capital-asset ratio then its efficiency will increase. But it may be an indication that higher capital ratios may prevent "moral hazard." As is often cited in discussions of the thrift crisis, as an institution's capital level decreases it has an increasing incentive to "bet the bank," since it stands to gain if the risk pays off and tends to lose only the amount of capital it has invested in the bank if the bet loses. Similarly, the managers of banks with lower capital levels might have more of an incentive to engage in perk-taking, and they face less shareholder scrutiny than banks with higher capital ratios. (If the owners' stake, that is, capital, is low, owners have less incentive to make sure the bank is run efficiently.[28]) Therefore, higher bank capital may not only provide a cushion for the deposit insurance fund, it might also provide appropriate incentives to bank managers to avoid waste. The capital-asset ratio might also be significantly related to inefficiency because inefficient banks have lower profits, which might lead to lower capital-asset ratios in the future.[29]

CONCLUSION

Banks in the Third District appear to be operating at cost-efficient output sizes and product mixes, but there appears to be a significant level of X-inefficiency at our banks. Some banks apparently are not using their labor, plant and equipment, and funds in the most efficient way possible, and case studies that focus on the more efficient banks in the District

[26]The simple correlation coefficient indicates that being located in New Jersey is significantly related to being inefficient, but this is because the New Jersey banks in the sample tend to have lower capital ratios than Pennsylvania and Delaware banks in the sample. Once capital ratio is controlled for (as in the regression), being located in New Jersey is not significantly related to inefficiency.

[27]There are a few other statistically significant relationships. For example, inefficient banks tend to have a higher percentage of their loans in construction and land development; national banks appear to be less efficient than state banks that are members of the Federal Reserve System but seem to have the same level of efficiency as state nonmember banks. (Note: all nationally chartered banks are Fed member banks, but their primary regulator is the Office of the Comptroller of the Currency, not the Fed.)

[28]Mester (1990) discusses the incentive effects of bank capital in mitigating bank risk-taking.

[29]But this is probably not the entire reason, since the relationship between capital assets and inefficiency holds even when return-on-assets is held constant, and while return-on-assets and capital assets are correlated, they are not collinear.

might shed light on how greater efficiency can be achieved. Theoretical advances may enable us to better identify the sources of the inefficiencies and verify that measured differences in inefficiency are true differences and do not result just from omitting factors that affect cost or misspecifying the cost function.

In terms of coping with the increased competitive pressures, inefficient banks in the Third District have more to fear from banks that are efficient producers than from banks that are producing a particular output volume or product mix. There is less to be gained in terms of cost savings from changing output size or mix than from using inputs more cost-effectively. Inefficient banks will have to get costs under control or else be prepared to be driven from an increasingly competitive marketplace.

SUGGESTED READINGS

Berger, Allen, and David Humphrey. "Measurement and Efficiency Issues in Commercial Banking," in Z. Griliches, ed., *Measurement Issues in the Service Sectors*. Chicago: National Bureau of Economic Research and Chicago University Press, 1992a.

Berger, Allen, and David Humphrey. "Megamergers in Banking and the Use of Cost Efficiency as an Antitrust Defense," *Antitrust Bulletin*, 37 (Fall 1992b), pp. 541-600.

Calem, Paul S. and Loretta J. Mester. "Search, Switching Costs, and the Stickiness of Credit Card Interest Rates," Working Paper 92-24/R, Federal Reserve Bank of Philadelphia, revised January 1993.

Calem, Paul S. "The Proconsumer Argument for Interstate Branching," this *Business Review* (May/June 1993), pp. 15-29.

Clark, Jeffrey A. "Economies of Scale and Scope at Depository Financial Institutions: A Review of the Literature," *Economic Review*, Federal Reserve Bank of Kansas City (September/October 1988), pp. 16-33.

Evanoff, Douglas D., and Philip R. Israilevich. "Productive Efficiency in Banking," *Economic Perspectives*, Federal Reserve Bank of Chicago, (July/August 1991), pp. 11-32.

Faulhaber, Gerald R. "Profitability and Bank Size: An Empirical Analysis," The Wharton School, University of Pennsylvania, mimeo, 1993.

Hughes, Joseph P., and Loretta J. Mester. "A Quality and Risk-Adjusted Cost Function for Banks: Evidence on the 'Too-Big-To-Fail' Doctrine," *Journal of Productivity Analysis*, 4 (September 1993), pp. 293-315.

Humphrey, David. "Costs and Scale Economies in Bank Intermediation," in R. Aspinwall and R. Eisenbeis, eds., *Handbook for Banking Strategy*. New York: Wiley and Sons, 1985.

Kraus, James R. "The Whole World Is Watching Asset Quality, Rate of Return," *American Banker*, July 29, 1993, p. 2A.

Mester, Loretta J. "Efficient Production of Financial Services: Scale and Scope Economies," this *Business Review* (January/February 1987), pp. 15-25.

SUGGESTED READINGS

Mester, Loretta J. "Owners Versus Managers: Who Controls the Bank?" this *Business Review* (May/June 1989), pp. 13-23.

Mester, Loretta J. "Curing Our Ailing Deposit-Insurance System," this *Business Review* (September/October 1990), pp. 13-24.

Mester, Loretta J. "Agency Costs Among Savings and Loans," *Journal of Financial Intermediation*, 3 (June 1991), pp. 257-78.

Mester, Loretta J. "Traditional and Nontraditional Banking: An Information-Theoretic Approach," *Journal of Banking and Finance*, 16 (June 1992), pp. 545-66.

Mester, Loretta J. "Efficiency of Banks in the Third Federal Reserve District," Federal Reserve Bank of Philadelphia Working Paper 94-1 (1994).

O'Hara, Terrence. "Interstate Branching Yet to Sound Alarms," *American Banker*, August 2, 1993.

Schmidt, Peter, and Robin C. Sickles. "Production Frontiers and Panel Data," *Journal of Business and Economic Statistics*, 2 (October 1984), pp. 367-74.

New Indexes Track the State Of the States

*Theodore M. Crone**

Have recessions lasted longer in Pennsylvania than in the nation? When did the most recent recession begin in New Jersey? Did Delaware avoid the recessions of the early 1980s altogether? Questions about how business cycles differ from state to state are raised frequently in the popular press and in business commentaries. The answers are seldom clear, but the questions are not idle ones. Some industries, such as construction and retail trade, are particularly sensitive to the local business cycle. Since people tend to live close to their jobs and shop close to where they live, sales of new homes, cars, and many consumer items depend on the prospects for jobs and income in the local area. If the local economy is weakening, these prospects are poor; if the economy is strengthening, the prospects are better. So knowledge about where the region is in the business cycle can be critical for managers in many businesses. But economic data are often ambiguous and sometimes contradictory; one indicator may be showing improvement while another shows decline. For example, the unemployment rate may be up at the same time job levels are increasing. Composite indexes, constructed from a number of individual indicators, can help clear up the ambiguity. There are

*Ted Crone is Assistant Vice President in charge of the Regional Economics section in the Philadelphia Fed's Research Department. Ted would like to thank James Stock and Mark Watson for a copy of the programs that estimate their new composite index of coincident indicators. He also thanks Keith Sill for help in adjusting Stock and Watson's programs to calculate the new state indexes.

commonly accepted composite indexes for the national economy, such as the index of leading indicators, but composite indexes are not readily available at the regional level, making it more difficult to track regional business cycles. This article introduces new composite indexes for the three states in the Third Federal Reserve District that make use of statistical techniques previously used for national indexes but not regional ones.

COMPOSITE INDEXES CAN HELP TRACK BUSINESS CYCLES

Between 1970 and 1990 real output in the U.S. grew at an average annual rate of 2.7 percent, and employment increased 2.2 percent a year. But output and employment fluctuated widely around these trends as the economy went through several business cycles. Economic trends vary from region to region, but all regions are affected by national business cycles, and some regions have exhibited cycles of their own.

Almost 50 years ago, Arthur Burns and Wesley Mitchell fashioned the commonly accepted description of a business cycle:

> *Business cycles are a type of fluctuation found in the aggregate economic activity of nations that organize their work mainly in business enterprises: a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals which merge into the expansion phase of the next cycle . . . ; in duration business cycles vary from more than one year to ten or twelve years; they are not divisible into shorter cycles . . . [that exhibit swings in economic activity of similar] amplitudes.*[1]

A complete cycle from one peak to the next consists of a recession or contraction followed by a recovery phase in which economic measures return to their previous peaks and then an expansion phase in which the measures reach new peaks. The description by Burns and Mitchell mentions three criteria for dating the various phases of the business cycle. The contractions and expansions must be broad-based, that is, they must occur in many sectors and be reflected in several indicators (*diffusion*). They must last a sufficient length of time (*duration*). And the change from peak to trough or from trough to peak must be sufficiently large (*amplitude*). All three criteria must be satisfied in order to define a contraction or expansion. A sharp decline in one sector, such as agriculture, would not qualify as a recession if it did not spill over into other sectors of the economy; the decline would not be broad enough. On the other hand, a broad-based decline that was very brief, one quarter, for example, probably would not qualify as a recession; it would be too short. Likewise, two quarters of 0.1 percent decline in output might not qualify as a recession; the decline would not be deep enough.

Since business cycles are broad-based, they tend to generate their own momentum. Downturns in the economy can be set in motion by a variety of factors, such as a sharp increase in the price of a major resource like oil or a sudden large reduction in government spending. Once a general downturn begins, firms begin to lay off workers. This loss of jobs as well as the uncertainty among people who are still employed leads consumers to cancel or postpone purchases, which results in more layoffs. Not all sectors of the economy are equally vulnerable to a downturn; consumers may be reluctant to delay seeing a doctor if they are ill, but they might readily put off the purchase of a new car. In general, manufacturing industries are more sensitive to business cycles than service industries. A downward spiral in the economy might be halted by a change in consumer expectations that raises confidence, by an increase in disposable income through a reduction in taxes,

[1] Arthur F. Burns and Wesley C. Mitchell, *Measuring Business Cycles* (National Bureau of Economic Research, 1946).

or by a rise in government spending for goods and services—any of which would increase demand. When this increased demand becomes broad enough it creates its own momentum toward further expansion. Thus, knowledge of whether the economy is entering a recession or beginning a recovery is important to local businessmen.

Dating business cycles using the Burns and Mitchell criteria is not always straightforward; it involves some personal judgment. In practice, the official dating of recessions and expansions is done by the Business Cycle Dating Committee of the National Bureau of Economic Research (NBER), which is composed of professional economists. To avoid any possibility of political considerations in setting these dates, none of these economists is a government official. The committee considers a number of economic indicators in dating business cycles. Composite indexes, however, play a special role in these decisions because they combine information from several sources to indicate the general state of the economy. They include not only data for the overall economy, such as employment and personal income, but also data from individual sectors, such as retail sales or industrial production.

The Department of Commerce publishes three composite indexes for the national economy—indexes of leading, lagging, and coincident indicators.[2] Of these three, the composite index of coincident indicators is the most important for dating business cycles. A good index of coincident indicators should decline at or near the beginning of recessions and should rise at or near the end. In the last 45 years there have been 18 business-cycle turning points in the U.S. economy. With four exceptions, the highest and lowest levels of the Commerce Department's index during each cycle have been within three months of the official dates of the business-cycle peaks and troughs.[3] Although the NBER dating committee considers the coincident index when it sets the dates for business cycles, it is not obligated to set the dates at or near the turning points of the index. Therefore, the close correspondence between the official dates and the Commerce Department's index suggests that the index is coincident with the business cycle.

The Commerce Department's Composite Index of Coincident Indicators is constructed from four monthly data series—the number of jobs in nonagricultural establishments, personal income (minus transfer payments) adjusted for inflation, the index of industrial production, and manufacturing and trade sales adjusted for inflation. Month-to-month percent changes are calculated for each of these series, and the changes are standardized based on the long-run average absolute monthly change in the series. For example, the average absolute percentage change in monthly employment between 1948 and 1985 was 0.32 percent. Thus, if the change in nonfarm employment were 0.64 percent this month, the standardized change for this indicator would be 2 (i.e., 0.64 / 0.32). A preliminary coincident index is formed based on the average of the standardized changes in the components that make up the index. To obtain the Department's official composite index, this preliminary index is adjusted to grow over time at the same rate as real gross national

[2]This article was prepared before the most recent revisions of the Commerce Department's composite indexes. All references in this article refer to the unrevised indexes. See George R. Green and Barry A. Beckman, "Business Cycle Indicators: Upcoming Revision of the Composite Indexes," *Survey of Current Business*, Vol. 73, 10 (Oct. 1993), pp. 44-51.

[3]As the name suggests, an index of leading indicators should peak several months before the economy goes into recession and should reach its cyclical low before the recession ends. The timing should be reversed for an index of lagging indicators. The leads and lags in the Commerce Department's leading and lagging indexes have varied from as long as 23 months to as short as 1 month.

product and is set to 100 in 1982.[4] Except for the adjustment to account for differences in the average monthly changes in the four indicators, each indicator is given the same weight in forming the composite index. But some indicators like the total number of jobs may better reflect the overall state of the economy than other indicators like manufacturing and trade sales. Thus, while the Commerce Department's index has tracked national business cycles very well, it has been criticized for not being derived from a formal mathematical or statistical model.[5]

In order to support the theory of business cycles and aid in the dating of recessions and expansions, James Stock of Harvard University and Mark Watson of Northwestern University have constructed a new index of coincident indicators.[6] Using time-series econometric techniques, they formalized the notion that the business cycle is best measured by the common movements across several economic data series. Each monthly indicator is thought of as having two components. The first is the general "state of the economy," which affects all the monthly indicators. It is not observed directly but only in the common movement of the indicators that are observed. The second component is an idiosyncratic element that might cause any one indicator to move in ways not associated with the general state of the economy. Stock and Watson's coincident index is an estimate of the common component. The movement of this unobserved state of the economy is reflected in varying degrees in each of the published monthly series used to estimate the composite index. Moreover, for some series, changes in the general economy could be reflected not only in the current month but also in succeeding months, and for other series, changes in the general economy could be foreshadowed in preceding months (see *A Formal Model of the New Coincident Index*). In effect, the Stock and Watson index is a weighted average of current and past values of the individual indicators, with the weights determined by the degree of common movement in the indicators.

In constructing their coincident index Stock and Watson used the same data series as the Department of Commerce, with one exception: they substituted employee hours in nonagricultural establishments for the number of nonagricultural jobs because economic output depends not only on how many people are working but also on how long they work. Stock and Watson's new index is available from 1959, and over that period it has coincided with the official business cycles even more closely than has the Commerce Department's Index of Coincident Indicators. The cyclical highs and lows in the Stock and Watson index coincide exactly with the official business-cycle turning points except in 1969 when the new index peaks two months prior to the official turning point.[7]

CONSTRUCTING STATE INDEXES

The success of the Stock and Watson method in constructing a national coincident index that tracks the official business cycles so closely suggests that this method could be used successfully to construct an index for state economies. But the construction of a comparable

[4]See "Composite Indexes of Leading, Coincident, and Lagging Indicators," *Survey of Current Business*, U.S. Department of Commerce, November 1987.

[5]The entire attempt to define business cycles was criticized from the beginning as an exercise in measurement without theory. See Tjalling C. Koopmans, "Measurement Without Theory," *Review of Economics and Statistics*, 29 (1947), pp. 161-72.

[6]James H. Stock and Mark W. Watson, "New Indexes of Coincident and Leading Economic Indicators," *NBER Macroeconomics Annual* (1989), pp.351-94.

[7]Of course, in developing their index Stock and Watson were attempting to trace the official business cycles prior to 1990, and the NBER dating committee had the new Stock and Watson Index when it dated the most recent recession.

A Formal Model of the New Coincident Index

The basic notion that a change in a monthly indicator reflects a change in the underlying state of the economy is captured in the following equation:

$$\Delta I_t = a + b\,\Delta S_t + u_t \quad (1)$$

where:

ΔI_t = the change in the observed monthly indicator between time t-1 and time t, and

ΔS_t = the change in the unobserved state of the economy between time t-1 and time t.[a]

Since the purpose of this model is to form a composite index, this equation is applied to a number of monthly indicators. For example, Stock and Watson use four monthly indicators so there are four equations similar to equation (1) in their model. The coefficients (a and b) will vary with each equation, but the unobserved variable (ΔS_t) is the same. In addition, the error term in equation (1) and the unobserved variable are assumed to follow an autoregressive process, so that

$$u_t = g_1 u_{t-1} + g_2 u_{t-2} + e_t \quad (2)$$

and

$$\Delta S_t = c + f_1 \Delta S_{t-1} + f_2 \Delta S_{t-2} + z_t \quad (3)$$

where e_t and z_t are error terms. Equations (2) and (3) are the transition equations in the system.

This system of equations (1) through (3) can be estimated using maximum likelihood techniques to produce an estimate of the change in the unobserved state of the economy (ΔS_t).[b] If we then index the unobserved variable S_t to equal 100 at some point in time, we can construct a time-series of the so-called "state of the economy," or a coincident index.[c]

---

[a]If the monthly indicator also reflects prior changes in the state of the economy, the estimating equation becomes $\Delta I_t = a + b_0 \Delta S_t + b_1 \Delta S_{t-1} \ldots + u_t$. If the monthly indicator partially foreshadows a change in the general state of the economy, the lagged values of the unobserved state of the economy are replaced by leads.

[b]In the actual estimating equations, Stock and Watson use the log difference of the monthly indicators. The change in the log of the monthly indicator is normalized by subtracting the historical mean and dividing by the standard deviation. Thus the constants a and c do not have to be estimated, and the unobserved variable that is estimated is the normalized change in the log of S_t.

[c]Stock and Watson set their national index at 100 in July 1967, and we set our state indexes at 100 in July 1987.

state index is not a simple matter of estimating Stock and Watson's model using state data. The monthly indicators used by Stock and Watson are not available at the state level. Moreover, there is no direct way to determine whether a composite index using other indicators at the state level would coincide with the business cycle because there are no official dates for state business cycles. Indeed, this was one reason for developing state indexes. To address the problem of finding an appropriate set of indicators to construct state indexes we identified a set of monthly indicators that are available at both the national and state levels. We selected those variables that were useful in dating national business cycles and assumed they would also be useful in identifying cycles in the state economies.

This indirect method resulted in identifying four variables to be used in our state indexes of monthly indicators—the total number of jobs in nonagricultural establishments, real retail sales, average weekly hours in manufacturing, and the unemployment rate.[8] These variables differ somewhat from those used in other national indexes. The total number of jobs in nonagricultural firms is used in the Commerce Department's Index of Coincident Indicators and was used in an earlier version of the Stock and Watson index.[9] The sales data used in our indexes are less comprehensive than those used by the Commerce Department and by Stock and Watson. Both use a series that includes sales by manufacturers and wholesalers as well as by retailers. Two of the variables we selected to construct our indexes, average hours in manufacturing and the unemployment rate, have not traditionally been counted among the coincident indicators. But we included them in our index because doing so improved the correspondence between the index and the official business cycles, compared to an index using only employment and retail sales.

Using these four variables we developed a national index and examined how closely it coincides with the official dates of national business cycles and with other composite indexes for the nation. Based on data since 1972, the pattern of the new national index follows closely the pattern of the Commerce Department's coincident index and Stock and Watson's coincident index (Figure 1).[10] There

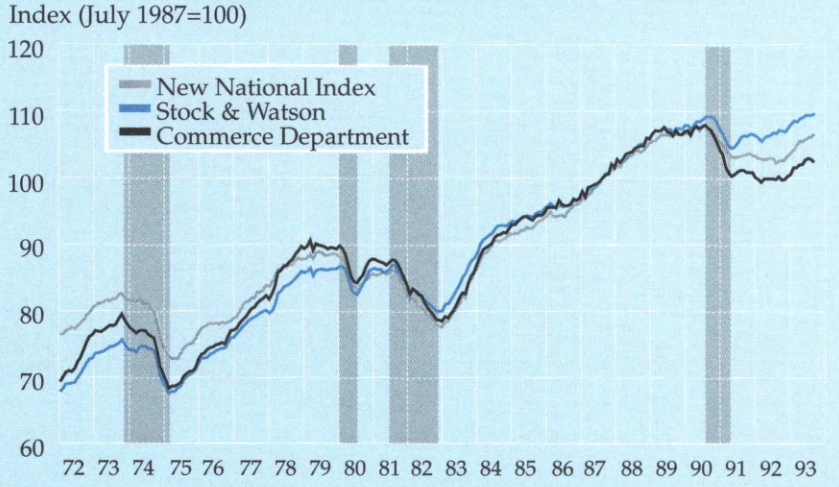
[8]The first three variables enter our model in log difference form. The unemployment rate enters in first difference form and is modeled to reflect the current value and three lags of the variable that reflects the state of the economy.

[9]James H. Stock and Mark W. Watson, "A Probability Model of the Coincident Economic Indicators," in Geoffrey Moore and K. Lahiri, eds., *The Leading Economic Indicators: New Approaches and Forecasting Records* (Cambridge University Press, 1990).

[10]We started the index in 1972 because some of the data series used in the index are not available at the state level prior to 1972.

FIGURE 1

Coincident Indicators—U.S.



The shaded areas represent the official recessions as determined by the NBER Dating Committee.

have been four national business cycles since 1972. Only the Stock and Watson index coincides precisely with the official peaks and troughs of all of them. But the new national index developed with data series that are also available at the state level traces the four national business cycles closely. With two exceptions the peaks and troughs of this new composite index are within one month of the official peaks and troughs of the U.S. business cycles since the early 1970s (Table 1).[11] The Commerce Department's Composite Index of Coincident Indicators was also off by several months at the same two turning points. Thus, the timing of the new index compares favorably with the timing of the Commerce Department's index, and it can be considered a coincident index.[12]

Using the same monthly indicators as in the new national index, we constructed coincident indexes for each of the three states in the Third Federal Reserve District — Pennsylvania, New Jersey, and Delaware (see *New National and State Indexes*). Since retail sales data are not available for Delaware, that state's index included only three of the four indicators.[13] These

[11]The two exceptions are the peak preceding the 1980 recession when the index led the economy by seven months and the trough of the most recent recession when the index lagged the economy by 15 months. Prior to the 1980 recession, the Stock and Watson index, the Commerce Department's index, and the new national index were basically flat for almost a year. Although the new index and the Commerce Department's index peaked several months before the beginning of that recession, they changed very little in the intervening months. After the official end of the most recent recession in March 1991, the cyclical lows for the Commerce Department's index and the new national index lagged by several months. The *steep declines* in the two indexes ended, however, about the same time as the official end of the recession, and the two indexes improved temporarily shortly after the official end of the recession.

[12]The average monthly increase in the new national index between 1972 and 1992 was 0.13 percent, compared with 0.16 percent for the Commerce Department's index and the 0.19 percent for the Stock and Watson index. The variance in the monthly change for the new index is also smaller than the variance for the other two indexes. The correlation between monthly changes in the new index and the Commerce Department's index is 0.54, and the correlation between the new index and the Stock and Watson index is 0.55. Both correlation coefficients are significantly different from 0 and from 1 at the 0.01 level. The correlation between the Commerce Department and the Stock and Watson indexes is considerably higher at 0.93, because with one minor exception these two indexes are constructed from the same monthly indicators.

[13]A national index constructed from the three variables used in the Delaware index tracks the national business cycles slightly less accurately than the new national index constructed from all four variables. In some cases, e.g., the 1981-82 recession, the timing of the peaks and troughs of the two national indexes are identical.

TABLE 1

Leads and Lags of the New National Index at Business Cycle Peaks and Troughs

(leads and lags in months)

BUSINESS CYCLE PEAKS	lead (+)/ lag (-)	BUSINESS CYCLE TROUGHS	lead (+)/ lag (-)
November 1973	0	March 1975	-1
January 1980	+7	July 1980	0
July 1981	0	November 1982	-1
July 1990	+1	March 1991	-15

New National and State Indexes

Except for Delaware there are four measurement equations in each system used to estimate the new national and state indexes:

(1) $\Delta emp_t = \beta_e \Delta S_t + u_{te}$

(2) $\Delta hrs_t = \beta_h \Delta S_t + u_{th}$

(3) $\Delta rs_t = \beta_r \Delta S_t + u_{tr}$

(4) $\Delta UR_t = \beta_{u0} \Delta S_t + \beta_{u1} \Delta S_{t-1} + \beta_{u2} \Delta S_{t-2} + \beta_{u3} \Delta S_{t-3} + u_{tu}$

where

Δemp = the standardized change in the log of nonfarm employment

Δhrs = the standardized change in the log of average hours worked in manufacturing

Δrs = the standardized change in the log of real retail sales

ΔUR = the standardized change in the unemployment rate.

Since retail sales are not available for Delaware, equation (3) is omitted in the system of equations for the Delaware index. Lagged values of the unobserved state of the economy are entered in the unemployment rate equation because including the lags produced a national index that coincided better with the official NBER recession dates. Moreover, the unemployment rate is often a lagging indicator reflecting the state of the economy in previous months. The estimated coefficients for each of the four systems is given in the following table:

Estimates of Coefficients Used to Construct Indexes of Coincident Indicators

	US INDEX	PA INDEX	NJ INDEX	DE INDEX
EMPLOYMENT EQ				
β_e	0.715	0.530	0.823	0.701
	(.051)	(.081)	(.065)	(.134)
HOURS EQ				
β_h	0.175	0.175	0.159	0.185
	(.032)	(.041)	(.050)	(.053)
RETAIL SALES EQ				
β_r	0.156	0.128	0.046	
	(.026)	(.034)	(.049)	
UNEMPLOYMENT EQ				
β_{u0}	-0.428	-0.044	-0.202	-0.637
	(.052)	(.092)	(.058)	(.130)
β_{u1}	-0.213	-0.240	-0.102	-0.136
	(.052)	(.102)	(.058)	(.088)
β_{u2}	0.033	-0.161	-0.003	0.071
	(.045)	(.103)	(.061)	(.089)
β_{u3}	0.026	0.217	-0.010	-0.010
	(.045)	(.110)	(.055)	(.063)

( ) = standard error of the estimate

models produced estimates of an unobserved "state of the economy," or a coincident index, for each of the three states.

BUSINESS CYCLES IN THE STATES

The coincident indexes for the three states in the Third Federal Reserve District define business cycles that correspond generally to the four national business cycles since 1972. But the cycles in each state have differed in their timing and duration. These differences can be seen by comparing the peaks and troughs of the state indexes with the official NBER dates and with the peaks and troughs of the new national index (Figures 2 through 4). Since there are clear differences between state and national business cycles, we need to apply some criterion to the new indexes to identify recessions and expansions at the state level. The experience of the NBER dating committee illustrates that there is no simple rule that will always identify peaks and troughs in the business cycle, but there should be some minimum decline in the index in order to characterize a given period as a recession. We found that a cumulative decline four times the average absolute monthly change in the index clearly defined four recessions in the new national index since 1972, and these recessions corresponded closely with the four officially recognized national recessions over that time period. We used the same rule of thumb to identify recessions at the state level. The peak of the cycle can be dated by the high point in the index just prior to the cumulative decline. Likewise, the trough of a cycle can be dated by the low point in the index prior to a cumulative increase that is four times the average absolute monthly change.[14] This identification of recessions at the state level allows us to compare cycles in the Third District states to national cycles. The peaks and troughs of the state indexes are shown in Table 2.

[14]Other simple rules could be used to date the beginning and end of a recession, such as three or four consecutive decreases or increases in the index. While the use of such consecutive decrease or increase rules would move the peak or trough closer to the NBER date for some recessions, for other recessions they would move the peak or trough further away from the official dates. These rules were not clearly superior to using the absolute high point and low point of the index as the business cycle turning points, and they have no compelling theoretical justification.

TABLE 2
Peaks and Troughs of State Indexes

	PA INDEX	NJ INDEX	DE INDEX
PEAK	November 1973	November 1973	February 1973
TROUGH	May 1975	May 1975	April 1975
PEAK	June 1979	February 1980	February 1980
TROUGH	September 1980	July 1980	April 1980
PEAK	March 1981	September 1981	July 1981
TROUGH	February 1983	November 1982	January 1982
PEAK	March 1990	March 1989	June 1990
TROUGH	July 1991	September 1992	April 1991

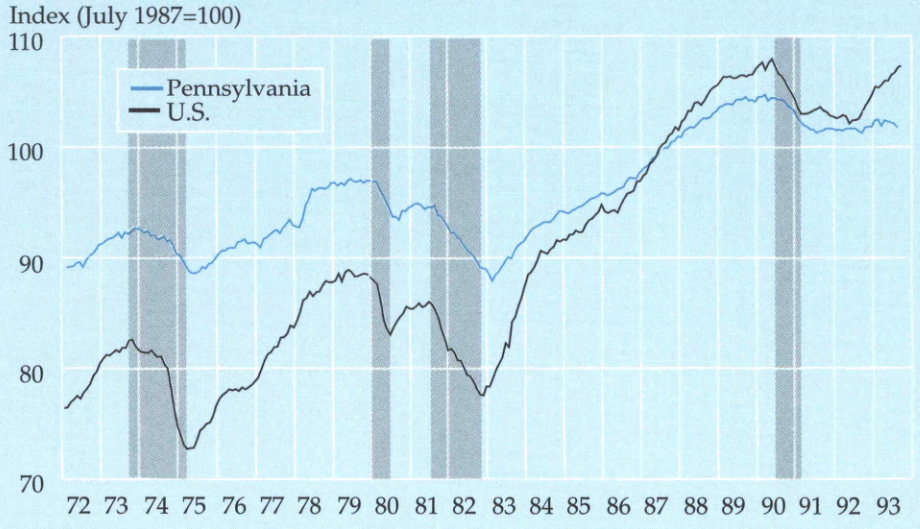
Pennsylvania's economy comprises somewhat less than 5 percent of the U.S. economy as measured by the number of jobs in the state and by gross product. In terms of the mix of industries, the cyclically sensitive manufacturing sector represents a larger percentage of the Pennsylvania economy than it does of the national economy, so one might expect Pennsylvania to suffer more recessions or longer recessions than the nation.[15] While there have not been more recessions in the state, recessions have generally lasted longer in Pennsylvania than in the nation (Figure 2). In every recession since 1972 the new coincident index for Pennsylvania has recorded a longer downturn than indicated by the official dates for the national recession. And except for the last recession, the declines in the Pennsylvania index have also lasted longer than the declines in the comparable national index. And generally recoveries in Pennsylvania have been less vigorous than in the nation as a whole. The current recovery is a striking example. At the end of the 1990-91 recession the Pennsylvania index technically reached its cyclical low 11 months before the new national index, but the state's index was little changed for more than a year after reaching that low point and was not signaling a recovery. The index reflected the popular impression of a lingering recession in the state.

New Jersey's economy represents slightly more than 3 percent of the U.S. economy in terms of jobs and gross product. The structure of the New Jersey economy has changed over the past 20 years from a greater than average dependence on manufacturing to a less than average dependence. Financial and business

[15]Over the period 1972 to 1992 manufacturing employment has averaged 25 percent of Pennsylvania's total employment but only 21 percent of U.S. employment.

FIGURE 2

Index of Monthly Indicators—PA



The shaded areas represent the official recessions as determined by the NBER Dating Committee.
The U.S. index is the new national index constructed from variables also available at the state level.

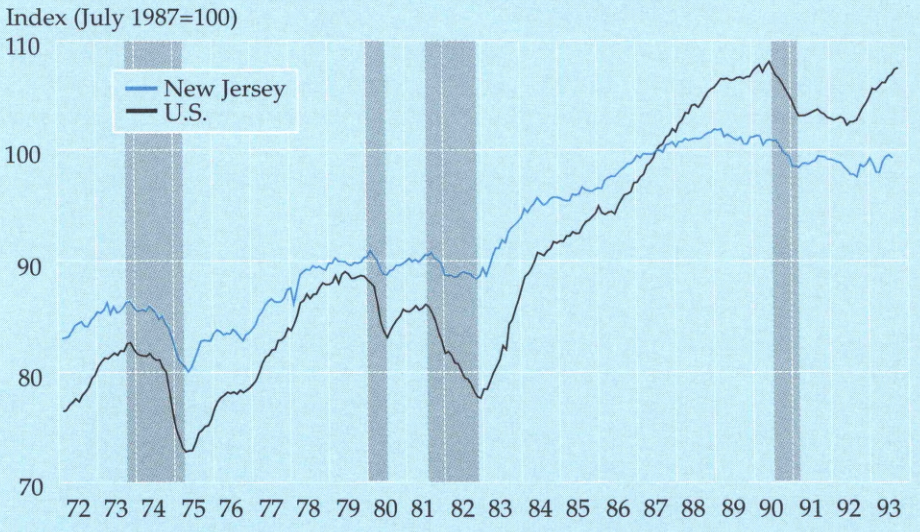
services have become a more important part of the state's economy. According to the new coincident index for New Jersey, some recessions in the state have been longer than the U.S. average (1973-75, 1990-91), and some have been shorter (1980, 1981-82). That pattern holds whether we measure national recessions by the official NBER dates or by the comparable national index (Figure 3). The most recent recession in New Jersey has been especially prolonged in part because this recession affected the service-producing sectors more than previous ones. Based on the peak and trough in the state's coincident index, the latest recession in New Jersey lasted from early 1989 to mid-1992, much longer than it did in the other states of the Third Federal Reserve District. There were some temporary improvements in the index over this three-year period, but none of the improvements were strong enough to qualify as a recovery.[16]

Delaware's economy is less than one-half of 1 percent of the U.S. economy in terms of jobs in the state and in terms of gross state product. The state's economy is more heavily concentrated in manufacturing than the U.S. economy, a fact that should tend to make it more cyclical. The very rapid growth in financial and business services since the early 1980s, however, has helped the state weather the last few recessions relatively well. Clear counterparts to three of the four national recessions since 1972 are apparent in the history of Delaware's new coincident index (Figure 4). The one national recession that has no counterpart in the Delaware index is the short-lived one in 1980. The decline from peak to trough in Delaware's monthly

[16]That is, the total increase in the index during these temporary improvements did not equal four times the average monthly change.

FIGURE 3

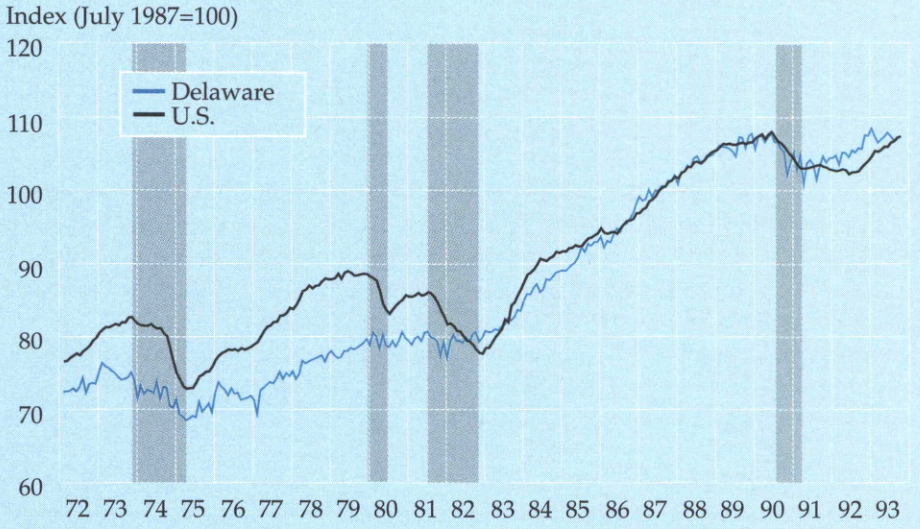
Index of Monthly Indicators—NJ



The shaded areas represent the official recessions as determined by the NBER Dating Committee. The U.S. index is the new national index constructed from variables also available at the state level.

FIGURE 4

Index of Monthly Indicators—DE



The shaded areas represent the official recessions as determined by the NBER Dating Committee.
The U.S. index is the new national index constructed from variables also available at the state level for many states.

index in the first half of 1980 was very brief (two months), and the total change in the index in those two months was less than three times the average monthly change. Based on the normal criteria for national recessions the brief 1980 downturn in Delaware would not qualify as a recession. The subsequent recession in 1981-82 is clearly discernible in the Delaware index, which registered a cumulative decline well over four times the monthly average, but this recession ended much earlier in the state than in the nation. Passage of legislation in 1981 encouraging the establishment of credit card banks in the state aided Delaware's economy. While the new index indicates that Delaware weathered recessions much better than the nation in the 1980s, it also indicates that the state suffered more in the 1970s. The 1973-75 recession began much earlier in Delaware than in the nation or in the other two states in the Third District. Moreover, the new coincident index suggests that Delaware suffered a local recession between February 1976 and February 1977—a downturn not matched at the national level. The state index declined a total of 5.9 percent, more than six times the average monthly change. The weakness in the state's economy was concentrated in the manufacturing and construction industries.

GETTING ANSWERS ABOUT STATE BUSINESS CYCLES

The ability to construct a composite index of monthly indicators that are available at the state level helps answer some of the questions frequently raised about regional business cycles. The new coincident index for Pennsylvania indicates that recessions have generally lasted longer in that state than in the nation as a whole. A set of indexes for all 50 states would undoubt-

edly uncover other states that tend to have longer recessions and help us identify some reasons. The new index for New Jersey indicates that the most recent downturn in that state began more than a year before the onset of the national recession and continued for more than a year after the end of the national recession. The index confirms that this recession was much longer in New Jersey than in the other states in the region. The new index for Delaware indicates that the state suffered only one recession in the early 1980s, and that one was briefer than the national downturn. But Delaware suffered a more extended recession than the nation in the early 1970s. Moreover, Delaware's index provides evidence of a local recession in the second half of the 1970s. The expansion of financial and business services in Delaware seems to have made the state's economy less cyclical.

These new composite indexes for the states provide another tool to monitor and analyze a region's economy. They can help us compare the timing of business cycles among the states and between any state and the nation. A full set of such indexes for all the states would help answer even more questions about regional business cycles and the structure of regional economies.

January/February
Leonard I. Nakamura, "Information Externalities: Why Lending May Sometimes Need a Jump Start"
D. Keith Sill, "Predicting Stock-Market Volatility"

March/April
Laurence Ball, "What Causes Inflation?"
Satyajit Chatterjee, "Leaning Against the Wind: Is There a Case for Seasonal Smoothing of Interest Rates?"

May/June
Edward G. Boehne, "Testimony on the Third District Economy and Monetary Policy"
Paul S. Calem, "The Proconsumer Argument for Interstate Branching"

July/August
Francis X. Diebold, "Are Long Expansions Followed by Short Contractions?"
Shaghil Ahmed, "Does Money Affect Output?"

September/October
B. Douglas Bernheim and *John Karl Scholz*, "Do Americans Save Too Little?"
Gerald A. Carlino, "Highways and Education: The Road to Productivity?"

November/December
Dean Croushore, "Introducing: The Survey of Professional Forecasters"
Laurence Ball, "How Costly Is Disinflation? The Historical Evidence"

FEDERAL
RESERVE BANK OF
PHILADELPHIA

Business Review Ten Independence Mall, Philadelphia, PA 19106-1574