Vol. 30, No. 2

ECONOMIC REVIEW

1994 Quarter 2

FEDERAL RESERVE BANK
OF CLEVELAND

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

ECONOMIC REVIEW

1994 Quarter 2
Vol. 30, No. 2

Economic Review is published quarterly by the Research Department of the Federal Reserve Bank of Cleveland. Copies of the *Review* are available through our Corporate Communications and Community Affairs Department. Call 1-800-543-3489, then immediately key in 1-5-3 on your touch-tone phone to reach the publication request option. If you prefer to fax your order, the number is 216-579-2477.

Coordinating Economist:
William T. Gavin

Advisory Board:
Ian Gale
Jagadeesh Gokhale
Joseph G. Haubrich

Editors: Tess Ferg
Robin Ratliff
Design: Michael Galka
Typography: Liz Hanna

Opinions stated in *Economic Review* are those of the authors and not necessarily those of the Federal Reserve Bank of Cleveland or of the Board of Governors of the Federal Reserve System.

Material may be reprinted provided that the source is credited. Please send copies of reprinted material to the editors.

ISSN 0013-0281

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

U.S. Banking Sector Trends: Assessing Disparities in Industry Performance

by Katherine A. Samolyk

Katherine A. Samolyk is an economist at the Federal Reserve Bank of Cleveland. The author thanks Robert B. Avery, Richard Dreese, Ian Gale, William Gavin, and James Thomson for helpful comments.

Introduction

The U.S. banking industry has a long tradition of decentralization as measured by geographic market structure. This feature largely reflects the impact of both inter- and intrastate branching restrictions as well as regulatory policies toward mergers and acquisitions. As a result of these policies, the industry comprises many small banks that operate in relatively localized and structurally diverse markets. In states allowing branching, banks tend to be fewer but larger than in unit banking states.

The phenomenon of bank holding companies emerged in the 1950s and 1960s as a response to restrictions on the scale and scope of banking activities. By holding banks as affiliates, a holding company can expand the geographic scale of its banking operations and broaden the scope of its nonbank activities to certain permissible lines of financial services. During the 1970s and 1980s, both the number of bank holding companies and the share of banks so affiliated increased, partly as a response to regulatory changes (Savage [1982], Amel and Jacowski [1989]). However, the trend also reflects changes in the environment in which these organizations operate.

While the U.S. banking industry has been consolidating into holding companies, it also appears to be shrinking. Domestic nonfinancial-sector debt grew substantially faster than GDP in the past decade, but the share of intermediated funds advanced by banks fell from 50 percent to 36 percent. The number of banks contracted by nearly 21 percent, from more than 14,400 in 1982 to about 11,400 at the end of 1992. The decade also witnessed a dramatic rise in bank failures and a spate of asset quality problems that translated into low industry profitability.

In assessing these trends, analysts have devoted considerable attention to the regional nature of the banking industry. Disparities in bank profitability over the past decade have been widely attributed to differences in local economic fortunes. Bank failures were largely concentrated in states experiencing economic difficulties. More recently, the poor performance of banks in New England and California has been associated with the so-called bicoastal recession.

Regional banking conditions also reflect the structural diversity across state banking sectors. Historically, bank failures have tended to occur in unit banking states, whereas institutions in branch banking states seem to have fared better

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

during periods of economic adversity.[1] In the past several years, however, the problems concentrated in large banks have raised concerns that a "too big to fail" regulatory policy is encouraging excessive risk-taking.[2]

Differences in economic fortunes and in bank structure across states complicate the assessment of industry performance. Are certain types of banks performing poorly because they are inherently different from other types of banking organizations, or do they happen to be concentrated in regions where the local economy is faltering? In this paper, I exploit the differences both within and across states in an attempt to evaluate how these factors were related to banking sector performance during the past decade. *Performance* refers to standard measures of banking conditions, including bank profitability, asset quality, capitalization, and lending. Data are compiled from individual Federal Financial Institutions Examination Council's *Reports of Condition and Income* (call reports) for each year between 1984 and 1992. I disaggregate state-level balance sheets and income statements to construct performance measures for banks that differ in size as well as in their holding company relationships. Then, controlling for state-specific economic factors, I examine the extent to which disparities in performance have been associated with differences in these bank characteristics.

The tone of this analysis is descriptive; the parsimonious number of relationships examined precludes a more causal interpretation. The findings reveal that the health of the local economy is indeed important in assessing the performance of the local banking sector. However, differences in banking conditions also appear to be associated with bank size and holding company affiliation; moreover, the emergent relationships are consistent with microeconomic studies that examine individual bank performance (Berger, Hanweck, and Humphrey [1987]). The results indicate that, controlling for local economic factors, relatively small banks (assets between $100 million and $1 billion in 1987 dollars) turned in the best performance over the past decade. In addition, smaller institutions that were affiliated with multibank holding companies had fewer problems with asset quality than did other small banks. Alternatively, the largest institutions — almost all of which belong to multibank holding companies — were less profitable. Hence, although banking fortunes reflect those of the local economy, performance also appears to be related to local industry structure. Judging at least by the experience of the 1980s, it seems that banks *can* be too large.

I. Industry Structure and Performance: An Overview

Assessments of the banking industry frequently describe it as a composite of banks that differ in size and location. For example, the *FDIC Quarterly Bulletin* presents industry data on banks classified by size and geographic region.[3] This focus reflects the view that such factors are important determinants of banking conditions.

The term *bank structure* is frequently used when referring to the characteristics of banking markets as well as those of individual institutions. Individual bank characteristics, such as the scale and scope of operations, can affect the costs at which banks produce financial services; hence the rationale for the focus on bank size. Market structure, measured by the relative size and number of firms, can influence the degree of local competition and, by extension, the quality, quantity, and price of financial services ultimately available to bank customers.

Researchers have studied how both market structure and individual bank characteristics are related to bank performance. One genre of studies looks at how market concentration is related to bank profitability and to the customer's cost of banking services.[4] Most findings reveal a positive relationship between market concentration and bank profitability. This result has been cited as evidence that more concentrated markets are less competitive. However, it also has been interpreted as an indication that more efficient firms tend to dominate the marketplace. A second line of research looks at how the costs associated with producing financial services are related to a

■ 1 This trend is less true in recent years. With the exception of Texas, failures during the past decade were not disproportionately located in unit banking states. Moreover, Wheelock (1993) notes that the choice of unit banking restrictions was popular in states with relatively cyclical economies, such as agricultural states.

■ 2 See Boyd and Gertler (1993) for a recent evaluation of this perspective.

■ 3 Similarly, assessments of changes in the structure of banking markets focus on how the geographic distribution of banks and the attendant concentration of banking markets have evolved (Amel and Jacowski [1989]).

■ 4 For example, see Berger and Hannan (1989).

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

BOX 1: DEFINITIONS

Structural Characteristics of Bank Cohorts

Size Classes (1987 dollars)

Very small: Less than $100 million in assets.
Small: $100 million to $500 million in assets.
Medium: $500 million to $1 billion in assets.
Large: $1 billion to $10 billion in assets.
Very large: More than $10 billion in assets.

Holding Company Affiliations

MBHC: Bank holding company holding more than one bank.
SBHC: Bank holding company holding only a single bank.
Independent: Not affiliated with a bank holding company.

Performance Measures of Industry Conditions

Lending and Capitalization

Capitalization: Bank equity capital as a percentage of total assets.
C&I lending: Commercial and industrial loans as a percentage of total assets.
CRE lending: Commercial real estate loans as a percentage of total assets.
Total bank lending: Total loans (including C&I loans, CRE loans, home mortgages, consumer loans, and other loans) as a percentage of total assets.

Bank Profitability and Asset Quality

ROA: Return on assets as measured by the ratio of annual net income to total assets.
Nonperforming assets: Past due loans (more than 90 days) plus nonaccruing assets plus other real estate owned, as a percentage of total assets.
Net loan charge-offs: The ratio of annual net charge-offs for loan losses to total bank loans as defined above.

NOTE: All measures use fourth-quarter data from the Federal Financial Institutions Examination Council's *Reports of Condition and Income* (call reports). Each performance measure is constructed from the cohort-level balance sheet or income statement. For example, ROA for each size class of banks is measured as the ratio of net income to total assets for each respective cohort of banks.

bank's structural characteristics.[5] Although the results are mixed, these cross-sectional assessments of bank efficiency have found evidence of modest economies of scale; the costs of providing banking services decline as firm size increases up to a relatively small size (Berger, Hanweck, and Humphrey [1987]).

The potential for the characteristics of banks and banking markets to affect industry performance motivates our interest in the phenomenon of bank holding companies. The importance of viewing holding company affiliation as a structural characteristic of banks depends on whether a bank in a holding company behaves differently than it would as an unaffiliated entity. At one extreme, holding companies may be passive vehicles that diversify across a number of banks and allow almost all decisions to be made at the subsidiary level. In this case, holding company affiliation might be unrelated to a bank's performance because it does not affect the bank's behavior. At the other extreme, if a bank can draw on its relationship with its holding company (for example, by reducing certain operating costs or increasing portfolio diversification through interbank loan sales), it may perform more like a larger institution.

Here, my focus on the link between the structural characteristics of banks and industry performance at the state level is more macroeconomic than microeconomic in nature. To the extent that banking conditions may impact credit availability, they may also affect economic activity. In a previous study using state-level data from the past decade (Samolyk [1992]), I found evidence suggesting that the health of the local banking sector plays a role in local economic fortunes. Banking conditions were more strongly related to current real personal income growth in states where the health of the banking sector was poor than in states where it was sound. Moreover, this relationship was not simply mirroring a correlation between banking conditions and past income growth. These findings suggest that if local bank characteristics affect local industry performance, they may have important economic consequences.

Both market factors and regulatory policies determine the structural characteristics of banks and banking markets. Here, I merely examine whether these characteristics have been associated with differences in banking sector performance. For example, small, localized banks may be more vulnerable to local economic distress, while larger banks are able to diversify over regional or even national markets. Thus, the state is not defined as the relevant "market" for banks of all types. Nevertheless, performance differentials across the various types of banking institutions within a state may provide evidence as to how bank characteristics can affect local banking conditions.

5 The scale of a bank's activities is usually defined in terms of balance sheet stocks, such as the volume of lending. The scope of a bank's activities refers to the composition of financial services it provides (for example, making loans versus funding securities). See Clark (1988) for a survey of these studies.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

TABLE 1

Banks and Banking Assets

Panel A	By Individual Bank Size		By Size of Largest Banking Organization	
	1984	1992	1984	1992
	Number of Banks			
Total	14,451	11,419	14,451	11,419
Very small	11,769	8,823	9,830	7,399
Small	2,171	2,037	1,985	2,196
Medium	210	229	412	344
Large	274	293	1,445	855
Very large	27	37	779	625
	Percentage of Banking Assets			
Total	100.0	100.0	100.0	100.0
Very small	15.2	11.6	12.0	9.2
Small	15.1	13.6	9.9	10.1
Medium	5.3	5.5	3.8	3.3
Large	28.7	32.1	27.2	18.8
Very large	35.7	37.2	47.1	58.6

Panel B	In Multibank Holding Company		In Single Bank Holding Company		Not in a Bank Holding Company	
	1984	1992	1984	1992	1984	1992
	Number of Banks					
Total	3,748	3,295	4,967	4,891	5,736	3,233
Very small	2,426	2,030	4,088	3,932	5,255	2,861
Small	989	871	728	828	454	338
Medium	132	141	58	64	20	24
Large	176	220	91	63	7	10
Very large	25	33	2	4	0	0
	Percentage of Banking Assets					
Total	67.0	72.7	23.2	20.7	9.8	6.6
Very small	3.7	3.0	5.5	5.3	6.0	3.3
Small	7.4	6.2	4.8	5.3	2.9	2.1
Medium	3.3	3.4	1.5	1.5	0.5	0.6
Large	18.7	26.2	9.6	5.3	0.4	0.6
Very large	33.9	33.9	1.8	3.3	0.0	0.0

SOURCE: Author's calculations.

II. Trends in Industry Structure

At the end of 1992, 11,419 domestic commercial banks filed call reports. Of these institutions, 71.7 percent were affiliates of a bank holding company. Of holding company affiliates, 40.3 percent were part of multibank holding companies and 59.7 percent were affiliates of single bank holding companies. Following convention, I characterize a bank's size in terms of the dollar value of its assets. Banks are placed in five size categories, which are adjusted for inflation so that a bank's classification will change only if its asset size has changed in real terms.[6] The inflation-adjusted (constant dollar) ranges for the five size cohorts are presented in box 1. Between 1984 and 1992, the unadjusted (current dollar) ranges of these size classes rose by approximately one-third.

Panel A of table 1 shows the distributions of banks and bank assets across the five categories. It also presents the distributions of banks and banking assets when each bank is classified by the size of its largest related organization. For example, in classifying a $100 million bank that is a subsidiary of a holding company with assets of $5 billion, I include that bank's data in the larger size cohort. This classification illustrates the distribution of banking assets by firm size when multibank holding companies are treated as branches of the holding company.[7] Panel A shows the trend toward fewer, larger banks in the industry: The number of very small banks has declined markedly. It also indicates that at the holding company level, the past decade has witnessed very large banking organizations growing into even larger ones.

Panel B of table 1 presents the distribution of banks and of banking assets classified by both size and holding company affiliation as defined in box 1. It indicates that the decline in the number of small and very small banks reflects decreases in both holding company affiliates and unaffiliated (independent) institutions. These data also underscore the emergence of the bank holding company as a prominent organizational entity. However, independent banks continue to be well represented, especially among smaller institutions.

Panel B of table 1 is also useful for understanding the measures employed in assessing banking conditions at both the national and

6 Both the *FDIC Quarterly Report* and the *Federal Reserve Bulletin* publish data on performance trends for banks classified by nonindexed size cohorts. Empirical studies that use cross-sectional data in a given year do not need to index nominal asset size classifications. However, studies that pool data on banks across time should deflate asset values into real terms to evaluate differences associated with bank size. For example, Avery and Berger (1991) index their classification of large and small banks in assessing the implications of risk-based capital on these segments of the industry.

7 The data on the number of banks require more clarification, as this category refers to the number of banks *affiliated with* holding companies of a given size, not the number of holding companies of that size. This indicates the potential misclassification associated with ignoring holding company affiliations.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

TABLE 2

Bank Balance Sheets: Lending and Capitalization (percentage of bank assets)

	All Banks	Very Small	Small	Medium	Large	Very Large
Bank lending						
1984	60.2	52.9	56.5	58.5	60.5	65.0
1989	62.4	52.9	60.0	65.2	65.4	63.4
1992	58.0	51.6	56.7	60.0	60.1	58.3
Commercial and industrial loans						
1984	22.6	13.1	16.8	18.9	21.0	30.9.
1989	18.8	10.7	13.9	16.9	18.8	23.5
1992	15.3	8.8	10.5	12.2	14.8	20.0
Commercial real estate loans						
1984	7.3	6.9	10.1	9.4	8.3	5.2
1989	11.3	9.2	13.5	14.8	13.2	8.8
1992	10.4	10.2	14.1	14.5	11.4	7.5
Capitalization						
1984	6.1	8.6	7.2	6.9	5.8	4.8
1989	6.2	8.9	7.6	6.7	6.1	4.8
1992	7.5	9.3	8.3	7.7	7.6	6.6

SOURCE: Author's calculations.

the state level. The focus here is on how performance differs across banks of various sizes and holding company affiliations. I construct national-level fourth-quarter performance measures for each of the five size classes (and each size class disaggregated by the three types of holding company relationships) by first aggregating the call report data for U.S. commercial banks in the size cohorts illustrated in panel B. The aggregated balance sheet and income statement of each cohort is then used to construct measures of capitalization, lending, bank profitability, and asset quality. These measures are defined in box 1. The state-level measures analyzed in the study are also constructed in this manner—albeit with the call report data on the individual banks for a given state. As a point of reference, I first examine the recent trends in banking conditions evident at the national level.

III. Trends in Bank Lending and Capitalization

Table 2 illustrates bank lending and capitalization (as percentages of bank assets) for the five size cohorts of commercial banks in selected years. Despite distinct differences in both loan/asset and capital/asset ratios across the size classes, these balance sheet measures have moved somewhat in concert during the past decade.

Larger banks appear to have invested a greater percentage of assets in loans than did smaller banks. While loan/asset ratios moved procyclically in medium and large banks, very small and very large banks did not exhibit this portfolio shift.[8] However, banks of all sizes changed the types of loans they funded over the period. The percentage of assets invested in commercial and industrial (C&I) loans declined and the percentage held as commercial real estate (CRE) loans rose in all segments of the industry. Large and very large banks moved most aggressively into CRE lending in the mid-1980s and have subsequently retrenched. For smaller banks, the monotonic shift to funding CRE loans is more indicative of a secular trend than of a cyclical real estate boom (and subsequent bust).

Larger banks also appear to have been less capitalized than smaller banks. However, the greatest disparities across the size classes occurred early in the 1980s, when industry capitalization was below 6 percent. Until the late 1970s, smaller banks faced higher capital requirements: They were viewed as riskier because they could not diversify as much as larger institutions. Subsequently, regulatory changes have eliminated differences in capital ratios based on size in favor of requirements associated with portfolio risk. The result has been that capitalization has increased in the industry as a whole, but by relatively more in larger banks.

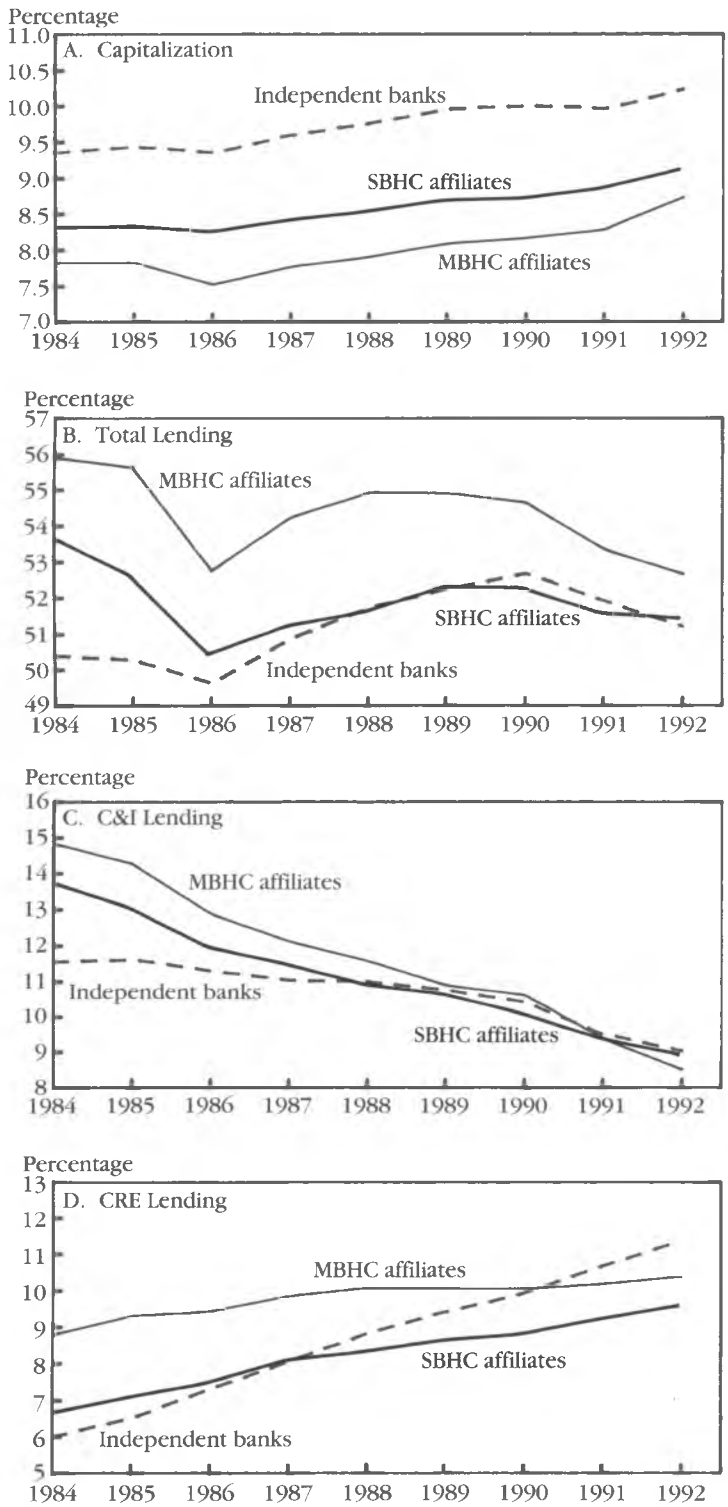
There also appear to be differences in bank lending and capitalization across similar-sized banks that vary in holding company affiliation. The panels in figure 1 depict capitalization and loan/asset ratios for very small banks sorted by their holding company relationships. Multibank holding company affiliates were less capitalized than otherwise affiliated institutions and made more loans than other very small banks,

8 See Boyd and Gertler (1993) for documentation of trends over the postwar period.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

FIGURE 1

Capitalization and Loan/Asset Ratios for Very Small Banks: by Holding Company Affiliation



SOURCE: Author's calculations.

although the differences in loan/asset ratios diminished over the decade. Thus, in terms of lending and capitalization, industry-level data suggest that multibank holding company affiliates behaved more like "larger" institutions than did other very small banks. This also appears to have been the case for small banks.

IV. Trends in Bank Profitability and Asset Quality

Differences in loan/asset ratios and bank capitalization are important factors in assessing the relative profitability and risk of banks. However, these variations do not inevitably translate into differences in risk or profitability. Although loans are a relatively risky class of investments (compared to securities), banks that have higher loan/asset ratios do not necessarily have riskier portfolios. Larger banks (or those affiliated with multibank holding companies) may be able to diversify the risks in their loan portfolios more successfully than smaller, more localized institutions. Moreover, larger banks may be profitable in spite of lower profit margins because their higher leverage allows them to pay a greater return to stockholders for a given return on their assets.

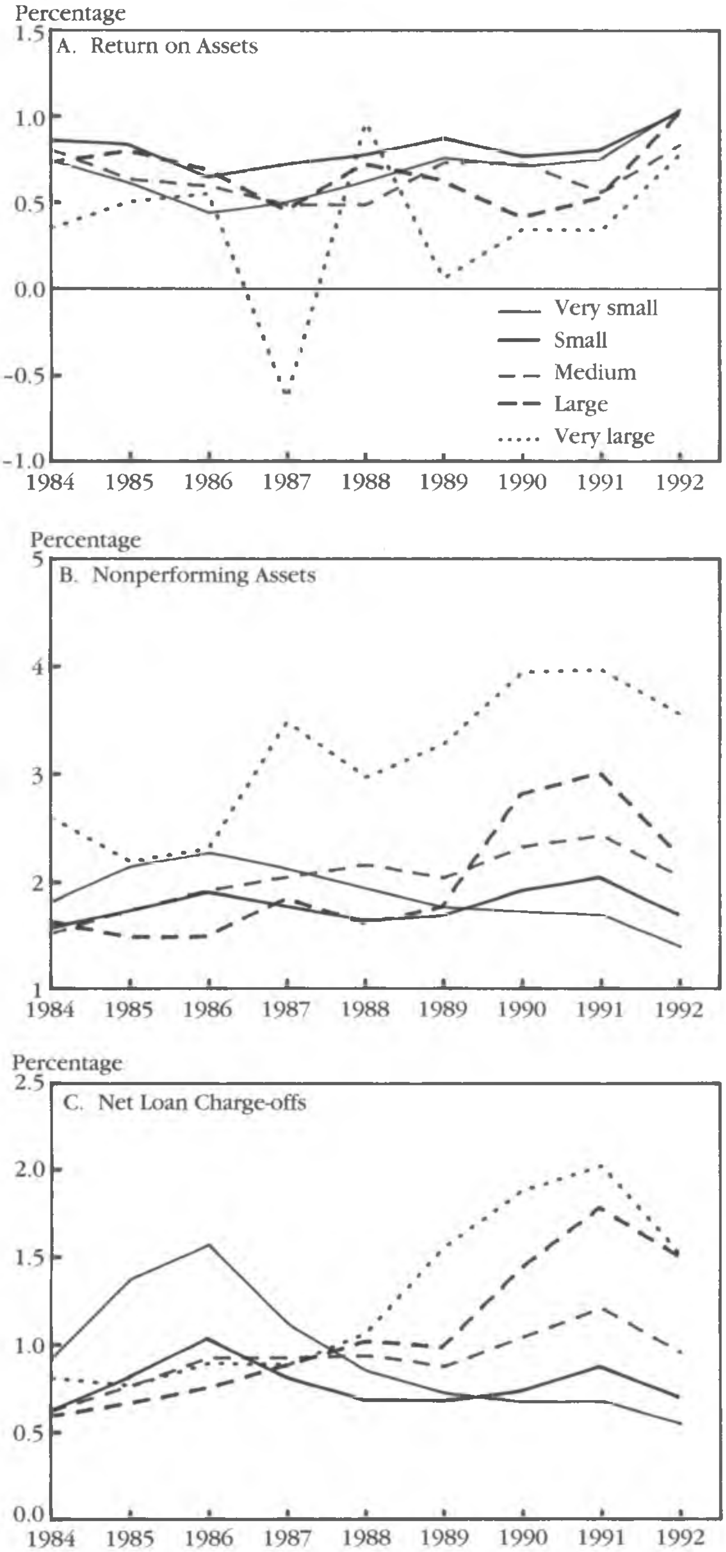
As evidenced by capital requirements in the past, smaller banks were viewed as riskier because of their limited ability to diversify. In the 1980s, government policies may have changed the relationship between bank size and bank risk by reducing the incentives for banks to manage losses prudently (Boyd and Graham [1991]). Deregulation, in tandem with changes in the treatment of problem institutions, may have increased the risks that uninsured investors allow banks — especially larger ones — to assume. The policy that banks can be too big to fail and the usual method of resolving bank failures (via purchase and assumption by healthy banks) shift the cost of bank failure from these investors to the Federal Deposit Insurance Corporation (and ultimately to taxpayers).

Unfortunately, ex ante portfolio risk and expected risk-adjusted yields are unobservable, so I examine data on ex post performance to infer indirectly how the risk–return relationship may vary across types of banks. I employ standard industry ratios used to measure bank profitability and asset quality (see box 1). Bank profitability is measured by the return on assets (ROA) for each class of banks. Problems with asset quality are measured by nonperforming

Digitized for FRASER
http://fraser.stlouisfed.org/ [redacted]
Federal Reserve Bank of St. Louis

FIGURE 2

Industry Performance Measures by Bank Size



SOURCE: Author's calculations.

assets as a percentage of total assets, and net charge-offs for loan losses as a percentage of total loans.

The panels in figure 2 illustrate how bank profitability and problems with asset quality varied across the five size classes of banks between 1984 and 1992. These performance measures reveal less of a discernible relationship to bank size and holding company affiliation than do bank lending and capitalization. Panel A indicates that the ROA for very large banks was more volatile than for smaller institutions. However, the differences in ROA do not indict size per se as an indicator of profitability. This is especially true from the perspective of bank stockholders; since larger banks are more leveraged, stockholders can earn a higher return on equity for a given ROA.

Panels B and C of figure 2 illustrate that smaller (primarily agricultural) banks experienced problems with asset quality in the mid-1980s. These problems have been widely attributed to the impact of local economic conditions. The dramatic rise in both nonperforming assets and loan charge-offs by larger institutions from 1988 to 1991 is commonly viewed as stemming from the troubled commercial real estate markets on the East and West coasts. These disparate economic conditions make it difficult to identify a consistent relationship between bank size and asset quality in the national-level data.

V. A Regional Perspective on Banking Sector Performance

Both banking sector performance and broader economic conditions varied widely across states during the past decade. At the same time, restrictions on branching and on bank holding company acquisitions were being eased in many states. In spite of these regulatory changes, a great deal of structural diversity remains both within and across state banking sectors. This diversity reflects the interaction of current regulatory environments with inherent market factors (such as size or population density).

Table 3 summarizes the differences in banking sector conditions across states in 1984 and 1992 in terms of the maximum, median, and minimum values of each measure as well as their means and standard deviations. The data mirror the trends evident at the national level, yet the variation across states is striking.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

TABLE 3

State-Level Commercial Banking Industry Ratios (percentage of total assets)

	Capitalization		CRE Lending		C&I Lending		Return on Assets		Nonperforming Assets		Net Loan Charge-offs	
	1984	1992	1984	1992	1984	1992	1984	1992	1984	1992	1984	1992
Maximum	9.7	12.8	22.0	22.5	30.3	24.4	1.6	3.0	3.6	7.0	2.0	5.1
Median	6.7	7.9	8.2	11.0	17.5	12.0	0.8	1.1	1.6	1.4	0.5	0.8
Minimum	5.2	6.5	1.6	2.4	7.2	5.7	−0.1	0.0	0.4	0.6	0.2	0.3
Mean	7.0	8.2	8.4	11.0	18.0	12.8	0.8	1.1	1.8	1.9	0.7	1.2
Standard deviation	1.0	1.3	3.5	3.9	4.5	4.1	0.3	0.5	0.8	1.3	0.5	1.0

NOTE: Net charge-offs are expressed as a percentage of total loans.
SOURCE: Author's calculations.

Table 4 presents the distribution of banks within and across states by their holding company affiliation in 1984 and 1992. Each state is ranked according to its total number of banks in 1984. Industry consolidation has been the rule rather than the exception in state banking sectors. The number of banks fell in 42 states during this period, and for 20 of these states the numbers dropped by more than 20 percent, although 38 states still had more than 50 banks at the end of 1992. These declines were accompanied by a decrease in the number of independent banks in 42 states. Overall, the percentage of smaller holding company affiliates also fell. This, however, is due to significant decreases in some states (most notably, Texas), which outweigh the increases in these affiliates in other states.

A trend toward one organizational type is not evident at the state level. Savage (1993) argues that the coexistence of holding company affiliates and independent banks within states indicates that there is not yet a dominant form of banking organization. The distribution of banks by size varies more substantially across states. In states with a tradition of unit banking, the industry tends to be populated by a large number of smaller institutions. As of year-end 1992, only 10 states had banks in the largest size cohort. Seven states had no banks with more than $1 billion (1987 dollars) in assets. However, except for very large banks, each size class is fairly well represented within and across states.

VI. Assessing Disparities in Industry Conditions

The diversity in the types of banks within and across states suggests a simple way of assessing the extent to which variance in bank performance can be attributed to differences in bank characteristics versus local economic conditions. In the following analysis, each state is treated as an individual sector composed of banks that vary in size and holding company affiliation. I then test for differences in industry profitability and asset quality that can be attributed to these structural characteristics, controlling for local economic conditions and other state-specific fixed effects.

The analysis features state-level data over the nine-year sample period of 1984 to 1992. As with the evaluation of national-level trends, I disaggregate state-level measures of industry conditions into cohort-level measures for the five bank size categories, crossed with the three types of holding company relationships. Hence, the annual data yield 15 potential observations in each year on industry conditions within a state. Not all states have banks in each class.[9] The nine years of data for 51 states (including the District of Columbia) yielded a data set of 4,062 observations on a given measure of bank performance.

The rationale for analyzing the performance of a cohort of banks rather than that of each

9 For example, almost all very large banks are multibank holding company affiliates (two are affiliates of single bank holding companies). The distribution of the data is discussed in the final section.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

TABLE 4

Distribution of Banks by Holding Company Affiliation

	Total Number		In Multibank Holding Company Percent		In Single Bank Holding Company Percent		Not in a Bank Holding Company Percent	
	1984	1992	1984	1992	1984	1992	1984	1992
Rhode Island	13	12	7.7	16.7	69.2	41.7	23.1	41.7
Alaska	15	8	20.0	37.5	33.3	37.5	46.7	25.0
Nevada	16	18	18.8	38.9	18.8	27.8	62.5	33.3
Washington, D.C.	19	22	5.3	40.9	52.6	31.8	42.1	27.3
Hawaii	21	17	23.8	23.5	4.8	11.8	71.4	64.7
Idaho	25	20	20.0	45.0	32.0	15.0	48.0	40.0
Maine	26	22	34.6	22.7	15.4	50.0	50.0	27.3
Vermont	27	21	22.2	28.6	33.3	42.9	44.4	28.6
Delaware	32	40	46.9	50.0	15.6	25.0	37.5	25.0
Arizona	46	38	2.2	42.1	32.6	31.6	65.2	26.3
Connecticut	50	48	16.0	18.8	20.0	29.2	64.0	52.1
New Hampshire	59	28	35.6	25.0	18.6	50.0	45.8	25.0
Utah	60	54	18.3	13.0	38.3	25.9	43.3	61.1
North Carolina	63	78	1.6	21.8	28.6	25.6	69.8	52.6
Oregon	72	48	13.9	20.8	20.8	27.1	65.3	52.1
South Carolina	73	81	1.4	18.5	28.8	38.3	69.9	43.2
Maryland	88	96	30.7	38.5	11.4	26.0	58.0	35.4
New Mexico	95	84	41.1	44.0	32.6	38.1	26.3	17.9
Washington	102	94	16.7	14.9	21.6	21.3	61.8	63.8
Wyoming	116	63	56.0	36.5	24.1	41.3	19.8	22.2
Massachusetts	124	63	48.4	19.0	23.4	46.0	28.2	34.9
New Jersey	125	106	38.4	23.6	20.8	37.7	40.8	38.7
South Dakota	140	121	23.6	28.9	32.9	47.1	43.6	24.0
Mississippi	153	121	0.0	5.0	46.4	53.7	53.6	41.3
Montana	167	120	44.9	35.8	29.3	40.8	25.7	23.3
Virginia	176	170	26.1	27.1	10.2	22.4	63.6	50.6
North Dakota	177	143	24.9	21.7	46.3	64.3	28.8	14.0
New York	190	177	30.0	19.8	21.1	46.9	48.9	33.3
West Virginia	227	164	18.5	59.8	16.3	18.9	65.2	21.3
Arkansas	258	259	16.7	34.0	39.9	44.4	43.4	21.6
Alabama	269	215	26.0	25.1	25.3	49.3	48.7	25.6
Tennessee	293	248	17.1	27.4	38.6	54.0	44.4	18.5
Louisiana	302	221	0.0	7.2	52.0	60.2	48.0	32.6
Ohio	320	271	31.3	32.1	12.2	26.6	56.6	41.3
Pennsylvania	326	281	15.3	33.1	22.7	40.6	62.0	26.3
Kentucky	336	311	4.5	37.3	49.7	46.9	45.8	15.8
Michigan	365	215	54.5	42.3	8.5	38.1	37.0	19.5
Indiana	378	270	2.1	43.3	46.3	39.3	51.6	17.4
Georgia	383	397	25.6	30.0	26.9	39.5	47.5	30.5
Florida	427	394	44.0	29.9	23.4	32.0	32.6	38.1
Colorado	446	349	49.6	47.6	30.3	29.5	20.2	22.9
California	449	451	6.9	8.9	34.7	31.9	58.4	59.2
Nebraska	472	374	10.0	26.5	61.4	52.9	28.6	20.6
Oklahoma	538	393	8.7	15.0	57.6	59.3	33.6	25.7
Wisconsin	590	445	32.0	40.2	28.3	35.7	39.7	24.0
Kansas	628	508	3.3	17.7	66.6	61.8	30.1	20.5
Iowa	629	542	21.8	29.7	49.4	54.8	28.8	15.5
Missouri	713	510	45.4	36.9	28.5	42.7	26.1	20.4
Minnesota	738	593	24.4	27.2	47.8	54.3	27.8	18.5
Illinois	1,241	1,006	25.9	37.0	33.1	40.6	41.0	22.5
Texas	1,853	1,089	40.9	20.7	23.0	40.8	36.0	38.6

SOURCE: Federal Financial Institutions Examination Council, *Reports of Condition and Income*, 1984 fourth quarter and 1992 fourth quarter.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

individual bank is to mitigate the effects of outliers and bank mergers. The cohort measures are averages of the individual banking data, where each bank is effectively weighted by its share of the cohort.[10] However, the findings should be similar to those obtained using individual bank data to assess differences across these classes of banks. Moreover, in examining the performance of classes of institutions, I can construct estimates of the performance differentials associated with bank size and holding company affiliation. These adjusted measures are directly comparable to the national-level data presented in figures 1 and 2.

Data on banks sorted by size and holding company affiliation within each state are used as cross-sectional observations on banking conditions in each year of the sample period. I then pool the data for each year to estimate reduced-form regressions for six measures of bank performance. To identify variance in performance that may be attributed to differences in bank size and holding company affiliation, I control for other factors that affect banking conditions both within each state and over time. In each regression, the following control variables are included: 1) a dummy variable identifying the state of an observation to control for state-specific differences in banking conditions during the sample period; 2) a dummy variable indicating the year of an observation to control for economywide variation in banking conditions over time; 3) the contemporaneous and lagged values of both the growth rate of state personal income and the volume of per capita failed business liabilities to control for the effect of local business conditions on banking sector performance; and 4) the ratio of state banking assets to state personal income in each year to control for variation in banking sector activity relative to that of the broader state economy.

Finally, to test whether bank size and holding company affiliation can explain differences in bank performance, I include dummy variables in each regression that measure intercept shifts for all classes of banks. Two different specifications for each measure of industry conditions are estimated; these vary the ways in which the dummy variables are interacted with time to examine whether performance differentials changed over the sample period. Both regression specifications include the same control variables and differ only in their treatment of the dummy variables.

VII. Evidence on Banking Sector Performance

Specification 1 includes an intercept shift for each size class (*Size*) and type of bank holding company affiliation (*HCA*) as well as the set of control variables. This specification takes the form

$$(1)\quad Ratio_{n,t,s,b} = \alpha + \beta_n^{State} + \beta_t^{Year} + \beta_S^{Size} + \beta_b^{HCA} + \sum_{i=1}^{5} \beta_i^E \, Econ_{i,n,t} + \varepsilon_{n,t,s,b},$$

where $Ratio_{n,t,s,b}$ is an observation of a bank performance measure, in the n^{th} state and t^{th} year, of banks in *Size* class s and *HCA* class b. $Econ_{i,n,t} (i = 1, ..., 5)$ is the set of state-level economic variables that includes current and lagged personal income growth, current and lagged per capita failed business liabilities, and the ratio of bank assets to personal income. Here, the intercept shifts associated with *Size* and *HCA* will measure how the average performance of banks with these characteristics varies. It is possible to estimate performance differentials only relative to a base group in each class. Very small banks are the *Size* base group and multibank holding company affiliates are the *HCA* base group.

Table 5 presents selected results of the regressions on six measures of banking conditions (summarized in table 3) obtained using specification 1. For brevity, the individual coefficient estimates of the intercept shifts for each state are not reported. However, they suggest that significant differences in banking conditions across states can be attributed to state-specific factors (other than current economic conditions) during the sample period. These may reflect average differences in local industry structure, including the structure of the banking sector. Similarly, the estimated intercept shifts for each year of the sample (relative to the base year, 1992) indicate that in evaluating bank performance over time, it is important to control for economywide trends that affect banks in all states. These coefficient estimates mirror the trends in banking conditions evident in the national-level data shown in

10 This method mitigates the effects of outliers within a class of banks. Outliers, in terms of a performance measure, will affect the measure only to the extent of their relative importance in the cohort. For example, to measure the ROA of small independent banks in each state (and each year), I take the ratio of their aggregated net income to their aggregated assets. A $50 million bank will, on average, contribute less to each term in the ratio than a $100 million bank. Of course, this is also the case for the data generally used in industry analyses.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

TABLE 5

Regressions Explaining Cohort-Level Performance

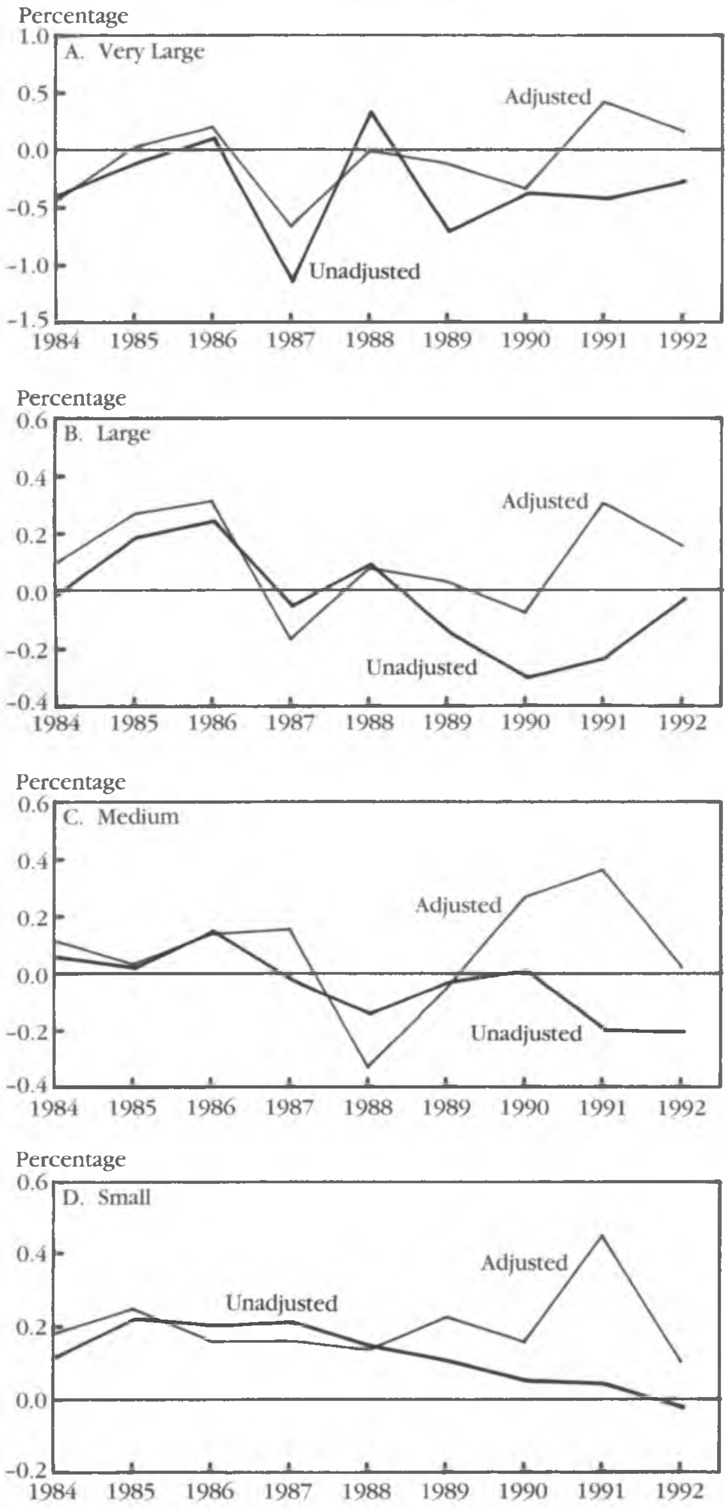
Dependent Variable	Capitalization	CRE Lending	C&I Lending	Return on Assets	Nonperforming Assets	Net Loan Charge-offs
R^2	0.146	0.407	0.397	0.164	0.285	0.188
Explanatory Variables						
State Dummies	(a)	(a)	(a)	(a)	(a)	(a)
Year Dummies	(a)	(a)	(a)	(a)	(a)	(a)
Economic Controls						
Personal income growth	−0.0155	−0.1816	−0.1464	0.07447	−0.2622	−0.0931
	(−0.37)	(−3.41)[a]	(−2.72)[a]	(6.65)[a]	(−15.69)[a]	(−7.21)[a]
Lagged personal income growth	0.0038	0.1043	0.0787	0.0661	−0.1780	−0.1071
	(0.09)	(1.99)[b]	(1.49)	(5.99)[a]	(−8.42)[a]	(−8.41)[a]
Failed business liabilities	−0.0010	−0.0025	−0.0003	−0.0017	0.0009	0.0010
	(−1.06)	(−2.09)[b]	(−1.16)	(−0.66)	(2.24)[b]	(3.33)[a]
Lagged failed business liabilities	−0.0018	−0.0027	−0.0009	−0.0009	0.0029	0.0014
	(−1.99)[b]	(−2.34)[a]	(−2.81)[a]	(−2.85)[a]	(7.59)[a]	(4.97)[a]
Bank assets to personal income	−0.0027	−0.0081	−0.0056	0.0053	0.001	0.003
	(−1.10)	(−2.66)[a]	(−1.81)	(8.25)[a]	(1.00)	(4.11)[a]
Holding Company Affiliation Dummies						
Not in a bank holding company	0.0070	−0.0258	−0.0272	0.0031	0.0022	0.0013
	(3.33)[a]	(−9.78)[a]	(−10.20)[a]	(5.65)[a]	(2.51)[a]	(1.99)[b]
In single bank holding company	−0.0062	−0.0084	0.0026	0.0012	0.0016	0.0005
	(−3.33)[a]	(−3.69)[a]	(1.13)	(2.53)[a]	(2.16)[b]	(0.89)
In multibank holding company	—	—	—	—	—	—
Size Dummies						
Very small	—	—	—	—	—	—
Small	−0.0173	0.0222	0.0150	0.0020	0.0003	−0.0009
	(−10.37)[a]	(10.54)[a]	(7.04)[a]	(4.49)[a]	(0.50)	(−1.77)
Medium	−0.0229	0.0188	0.0340	0.0008	0.0016	0.0009
	(−10.71)[a]	(6.99)[a]	(12.47)[a]	(1.34)	(1.80)	(1.34)
Large	−0.0304	0.0087	0.0432	0.0011	0.0021	0.0012
	(−14.42)[a]	(3.28)[a]	(16.10)[a]	(1.98)[b]	(2.38)[a]	(1.83)
Very large	−0.0389	−0.0276	0.0856	−0.0008	0.0066	0.0039
	(−9.93)[a]	(−5.60)[a]	(17.20)[a]	(−0.77)	(3.99)[a]	(3.23)[a]

a. Significant at the 1 percent level.
b. Significant at the 5 percent level.
NOTE: T-statistics are in parentheses. Coefficients of dummy variables indicate the intercept shift relative to the omitted category, as indicated by dashed lines.
SOURCE: Author's calculations.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

FIGURE 3

Adjusted and Unadjusted Performance Differentials in Return on Assets by Bank Size (Relative to Very Small Banks)



SOURCE: Author's calculations.

table 2 and figure 3 (again, for brevity, the individual coefficients are not reported).

Bank performance does appear to reflect local economic conditions, particularly in regard to bank profitability and asset quality. Both income growth and failed business liabilities help explain ROA, nonperforming assets, and loan charge-offs in the expected ways. Profitability as measured by bank ROA is positively related to income growth and negatively related to failed business liabilities. Symmetrically, asset problems measured in terms of both nonperforming assets and loan charge-offs are negatively related to income growth and positively related to failed business liabilities. C&I lending—and to a lesser extent CRE lending—is negatively related to failed business liabilities; banks appear to fund fewer loans when the credit quality of the local business sector deteriorates. However, the coefficients on state income growth suggest that end-of-year lending as a share of assets is also lower when recent income growth has been higher. The coefficients on the ratio of banking assets to state personal income are positive in the regressions explaining both ROA and loan charge-offs. Hence, when banking activity is high relative to economic activity, both bank profitability and problems with asset quality are higher as well. Finally, bank capitalization is relatively unrelated to the economic control variables.

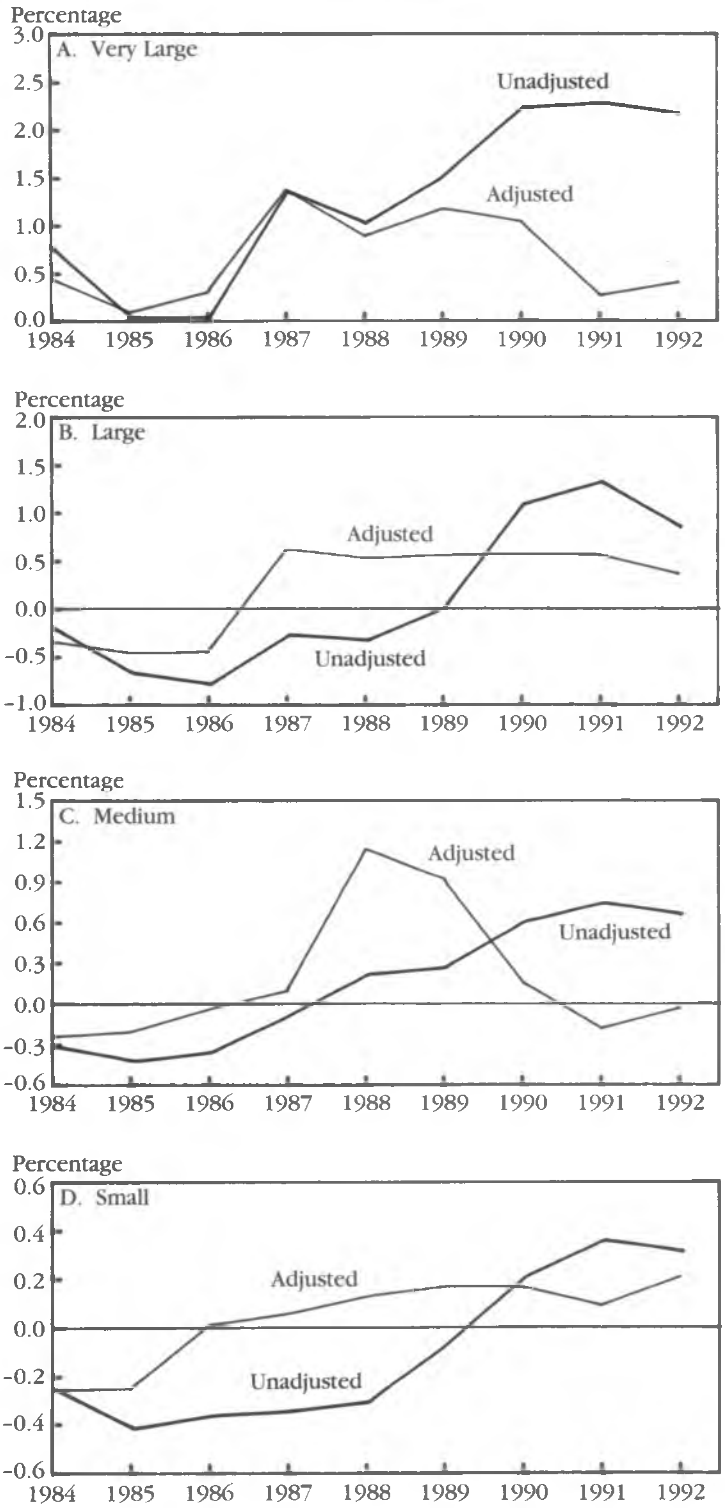
These findings, then, reveal that the profitability and asset quality of different segments of the industry to a large degree reflect the economic conditions impacting these institutions: When the local economy has been faring poorly, it is likely that the banking sector will follow suit. This analysis is consistent with most interpretations of banking trends. However, the results also hint that differences in bank performance can be attributed to differences in bank characteristics.

The results for specification 1 yield significant variations in banking conditions among institutions having different holding company relationships. Both single bank holding company affiliates and independent banks had a smaller share of assets invested in CRE loans than did multibank holding companies, while only independent banks held a significantly smaller share of C&I loans. Controlling for size, independent banks were, on average, more capitalized than multibank holding company affiliates, while single bank holding company affiliates were less capitalized. All else equal, both independent banks and single bank holding company affiliates earned a higher ROA

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

FIGURE 4

Adjusted and Unadjusted Performance Differentials in Nonperforming Assets by Bank Size (Relative to Very Small Banks)



SOURCE: Author's calculations.

than banks in multibank holding companies. However, independent banks and single bank holding company affiliates also had more problems on average with asset quality, as reflected in nonperforming assets and loan charge-offs.

Thus, controlling for size, the performance of banks in multibank holding companies differed from otherwise affiliated institutions; the former earned a lower average ROA, but also had fewer problems with asset quality. These affiliates invested a larger share of their portfolios in loans, but they appear to have been better risks than both single bank holding company affiliates and independent banks in terms of their performance during the past decade.

Bank performance also varied significantly across the five size classes. As illustrated in table 5, the differences in capitalization and in lending mirror those evident in the national-level data. The coefficients measuring the average differences in capitalization for each size class indicate that controlling for other factors, capitalization is inversely related to bank size. These coefficients imply that on average, the capital/asset ratio of very large banks was 3.9 percentage points lower than that of very small banks. C&I lending as a share of assets is positively related to bank size. Alternatively, while the middle three size classes of banks invested a higher percentage of assets in CRE loans than did very small banks, very large banks held a significantly smaller share.

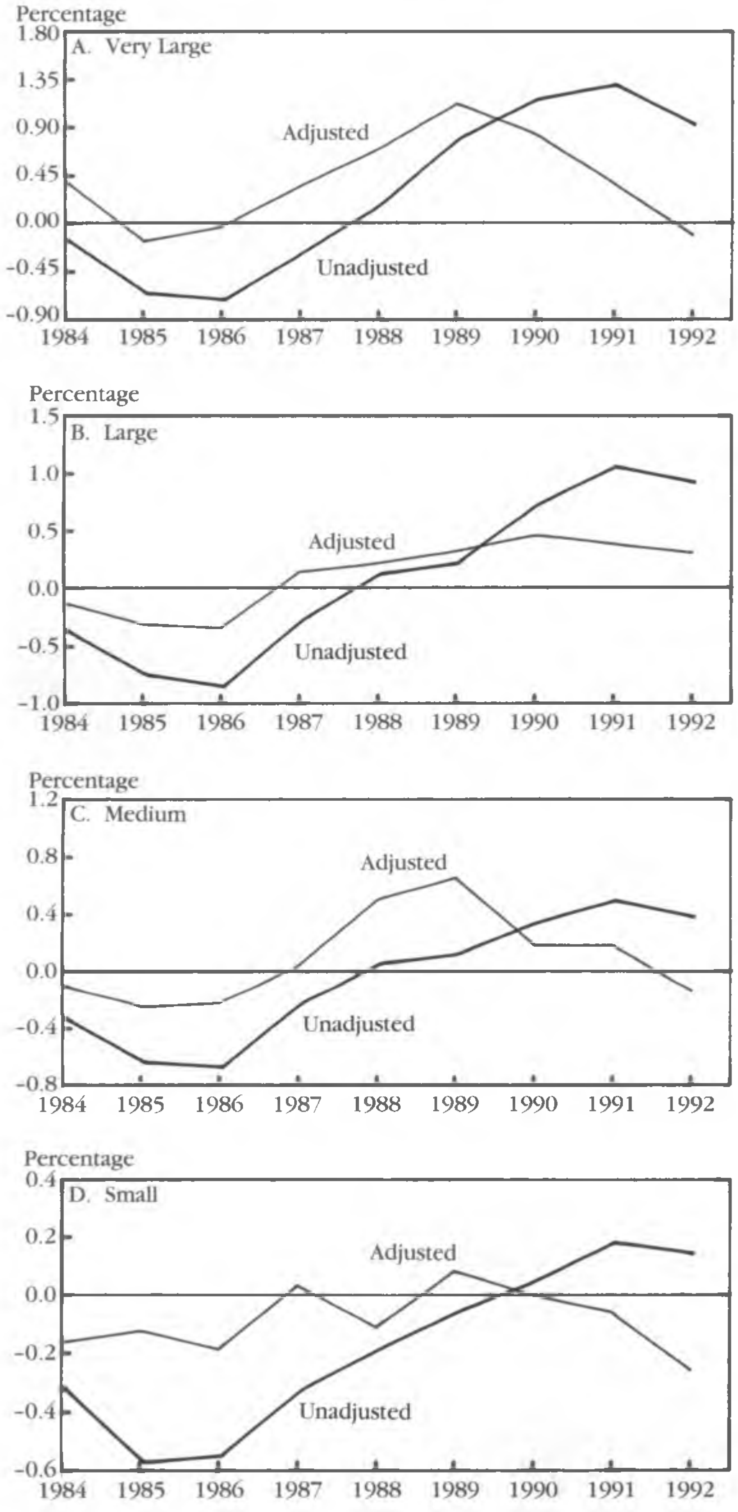
Perhaps not surprisingly, I find a less consistent relationship between bank size and bank performance in terms of profitability and asset quality. All else equal, small and large banks earned significantly higher ROAs than did either very small or very large banks. Large and very large banks had higher loan charge-offs and lower nonperforming asset ratios than did very small banks. Relatively small banks ($100 million to $500 million in assets, 1987 dollars) seem to have turned in the best performance in terms of profitability and asset quality. Interestingly, these banks are about the size that some studies have shown to maximize economies of scale (Berger, Hanweck, and Humphrey [1987]).

In summary, specification 1 estimates the average differences in bank performance that can be attributed to bank size and holding company relationships, controlling for local economic factors and aggregate trends that affect banking conditions. Thus, the results measure the extent to which the variation in bank performance *within* states is related to these structural characteristics. I find systematic

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

FIGURE 5

Adjusted and Unadjusted Performance Differentials in Net Loan Charge-offs by Bank Size (Relative to Very Small Banks)



SOURCE: Author's calculations.

differences in capitalization and lending across banks that vary in their size and holding company relationships. Indeed, the disparities in capitalization and lending observed at the national level appear to largely reflect these structural differences. The results also indicate some variation in bank profitability and asset quality across the different types of banks during the past decade.

VIII. Will Performance Differentials Continue?

The observed differences between large and small banks may have changed during the past decade. Estimating how the average performance of one type of bank (over a number of years) compares to that of another may obscure emerging differences in performance. For example, because of greater asset diversification, the asset quality of larger banks could have been significantly better than that of smaller banks early in the 1980s. If, indeed, these institutions have increased the relative risk of their investments (because they are too big to fail), they may still be on par with small banks in terms of average performance although their asset quality has been declining.

To examine the possibility that structural disparities in bank performance have changed in the past decade, I estimate a second set of regressions:

$$(2)\quad Ratio_{n,t,s,h} = \alpha + \beta_n^{State} + \beta_{s,t}^{Size*Year} + \beta_{h,t}^{HCA*Year} + \sum_{i=1}^{5} \beta_i^E \, Econ_{i,n,t} + \varepsilon_{n,t,s,h}.$$

Specification 2 includes an intercept shift for each *Size* and *HCA* class for *each year* in the sample period, as well as the control variables included in specification 1. Again, it is possible to estimate differences only relative to a base group in each class; hence the choice of very small banks as the *Size* base group and multibank holding company affiliates as the *HCA* base group. In these regressions, each intercept shift associated with a class of banks measures the estimated difference between the performance of that class and its respective base group *in a particular year*. Therefore, the estimated performance differentials associated with bank size and holding company relationships are allowed to vary over time.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

Here, I present the evidence of performance differentials in ROA, nonperforming assets, and loan charge-offs obtained using specification 2. The results for the control variables in these regressions are similar to those presented in table 5. Thus, I focus on patterns in the time-varying intercept shifts associated with bank size and holding company affiliation. The intercept shifts can be interpreted as annual bank performance differentials that have been adjusted for state-specific factors, local economic conditions, and economy-wide trends. These adjusted performance differentials are therefore estimates of the within-state variations in bank performance attributable to structural characteristics.

In figures 3, 4, and 5, the green lines illustrate the *adjusted performance differentials* in ROA, nonperforming assets, and loan charge-offs for each *Size* class of banks. It is instructive to compare these estimates to measured differences in performance that have not been adjusted for other economic factors. The national-level performance measures presented in table 2 are used to construct *unadjusted performance differentials* of this sort for each *Size* class of banks. For example, the unadjusted differential in the ROA of very large banks in each year is simply the ROA for this cohort of banks minus the ROA for very small banks (as illustrated in figure 2). The *unadjusted performance differentials* in ROA, nonperforming assets, and loan charge-offs are depicted by the blue lines in figures 3, 4, and 5.

The panels in figure 3 illustrate both the adjusted and unadjusted differences in ROA for each *Size* class of banks (relative to very small banks). A comparison of these series indicates that controlling for other economic factors mitigates the relatively poor performance of larger banks in recent years. The adjusted differentials in ROA for small banks suggest that, all else equal, they were more profitable than very small banks during the entire sample period. However, the adjusted differentials for the three largest classes reveal no emerging trends in profitability differentials that can be attributed to bank size per se.

The panels in figure 4 depict the unadjusted and adjusted differentials in nonperforming assets by *Size* class. The adjusted series do not exhibit the increasing disparities between sizes that are evident in the unadjusted data. Panels A and B do indicate that, all else equal, larger banks had significantly higher nonperforming asset ratios than did smaller banks in the late 1980s. Subsequently, however, the differences attributable to size decreased.

Finally, the panels in figure 5 illustrate both the adjusted and unadjusted differences in loan charge-offs for various-sized banks. The adjusted differentials are measurably smaller than the unadjusted ones. As in the case of nonperforming assets, controlling for other economic factors mitigates the emerging relationship between size and asset quality problems suggested by the unadjusted measures. However, again, it does appear that larger banks showed more asset quality problems than did smaller banks during the late 1980s.

The key result yielded by these series is that, controlling for economic factors, there is no evidence of a trend toward increasing disparities in bank performance that can be attributed to bank size as a structural characteristic. Similarly, the estimated performance differentials for the *HCA* classes do not suggest emerging disparities in bank profitability and asset quality associated with holding company relationships.

IX. Banking Sector Performance: Assessing the Trends

One interpretation of observed bank performance in the past decade is that the disparities between larger and smaller banks may indicate that increasing risks are being borne by the largest players (Boyd and Gertler [1993]). Accordingly, the trend toward larger banks has been viewed with some concern. This paper investigates the merits of this perception by descriptively assessing the extent to which differences in banking conditions can be attributed to variations in bank size and holding company affiliation.

Taking local economic factors into consideration reduces the disparities in bank performance attributable to these structural characteristics. There is some evidence that, after controlling for state-specific fixed effects, local economic conditions, and national-level trends, larger banks performed worse during the 1980s than did smaller institutions. However, the trend appears to have since reversed. In addition, the results indicate that, all else equal, banks that are associated with larger organizations through multibank holding company affiliations tended to perform better than otherwise-affiliated institutions.[11]

■ **11** In interpreting these findings, it is important to note that I am evaluating cohort-level banking conditions. Smaller banks that grow large because they are profitable are allowed to be reclassified into larger size cohorts. On the other hand, banks that are poor performers also may leave their cohorts as they are either closed or merged into larger institutions. Although beyond the scope of this study, it would be interesting to examine the extent to which trends in banking sector performance have been associated with changes in the population of banks across size classes and holding company affiliations.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

One potential explanation for the finding that large banks performed relatively poorly is that these institutions may be more likely to make loans outside their locality. To the extent that this is true, I do not control for economic conditions where they made loans. An obvious example is the case of the huge write-offs associated with loans to developing countries. The rationale for controlling for local economic conditions is that certain institutions are more subject to these conditions. Thus, evidence that less constrained firms are riskier may suggest that they could and did take on more risk during the past decade.

In assessing the performance of large banks, it is also important to note that most of these institutions are part of multibank holding companies. Thus, a study of the behavior of large banks is effectively a study of the joint effects of both large size and this form of banking organization. On the other hand, the evidence indicates that, all else equal, multibank holding company affiliations appear to benefit banks. This suggests that smaller affiliates have not experienced the same problems with asset quality as have larger institutions. Thus, it seems that the performance of larger banks reflects the effects of size rather than holding company status per se. Indeed, riskier loans may have been channeled to larger banks in the holding companies.

The results of this analysis indicate that the U.S. banking industry during the 1980s may have been characterized by a duality related to bank size. Although reregulation in the past several years has attempted to address this possibility, the evolving role of banks indicates that the link between bank size and bank performance merits further study from a regulatory perspective.

References

Amel, Dean F., and Michael J. Jacowski. "Trends in Banking Structure since the Mid-1970s," *Federal Reserve Bulletin*, vol. 75, no. 3 (March 1989), pp. 120 – 33.

Avery, Robert B., and Allen N. Berger. "Risk-based Capital and Deposit Insurance Reform," *Journal of Banking and Finance*, vol. 15, nos. 4/5 (September 1991), pp. 847 – 74.

Berger, Allen N., Gerald A. Hanweck, and David B. Humphrey. "Competitive Viability in Banking: Scale, Scope, and Product Mix Economies," *Journal of Monetary Economics*, vol. 20, no. 3 (December 1987), pp. 501 – 20.

Berger, Allen N., and Timothy H. Hannan. "The Price–Concentration Relationship in Banking," *Review of Economics and Statistics*, vol. 71, no. 2 (May 1989), pp. 291 – 99.

Boyd, John H., and Mark Gertler. "U.S. Commercial Banking: Trends, Cycles, and Policy," National Bureau of Economic Research Working Paper No. 4404, July 1993.

Boyd, John H., and Stanley L. Graham. "Investigating the Banking Consolidation Trend," Federal Reserve Bank of Minneapolis, *Quarterly Review*, vol. 15, no. 2 (Spring 1991), pp. 3 – 15.

Clark, Jeffrey A. "Economies of Scale and Scope at Depository Financial Institutions: A Review of the Literature," Federal Reserve Bank of Kansas City, *Economic Review*, vol. 73, no. 8 (September/October 1988), pp. 16 – 33.

Samolyk, Katherine A. "Bank Performance and Regional Economic Growth: Evidence of a Regional Credit Channel," Federal Reserve Bank of Cleveland, Working Paper 9204, February 1992.

Savage, Donald T. "Developments in Banking Structure, 1970–81," *Federal Reserve Bulletin*, vol. 68, no. 2 (February 1982), pp. 77 – 85.

________ . "Interstate Banking: A Status Report," *Federal Reserve Bulletin*, vol. 79, no. 12 (December 1993), pp. 1075 – 89.

Wheelock, David C. "Is the Banking Industry in Decline? Recent Trends and Future Prospects from a Historical Perspective," Federal Reserve Bank of St. Louis, *Review*, vol. 75, no. 5 (September/October 1993), pp. 3 – 22.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

Competition for Scarce Inputs: The Case of Airport Takeoff and Landing Slots

by Ian Gale

Ian Gale is an economic advisor at the Federal Reserve Bank of Cleveland. The author thanks Mark Stegeman for helpful discussions.

Introduction

The process of deregulating airlines in the United States began in earnest with the passage of the Airline Deregulation Act of 1978. The Act set in place a timetable for removing government controls on fares and entry into routes. However, airline access to airports has not been fully deregulated. In 1968, the Federal Aviation Administration (FAA) adopted the so-called "high-density rule" to combat congestion at four airports. Specifically, limits were placed on the number of operations per hour at Kennedy, La Guardia, O'Hare, and Washington National. More and more airports are becoming crowded, given the difficulties in securing permission from local authorities to expand existing facilities or to build new ones, so this problem will persist.[1]

For many years, the right to take off and land at the four crowded airports was determined by a committee system. Multilateral negotiations took place among the incumbent carriers and prospective entrants, with the FAA stepping in if an impasse occurred. Since 1986, incumbents have had property rights called "takeoff and landing slots" or simply "slots." A slot permits the owner to make one operation (a takeoff or landing) during a specified time period. Carriers now have the right to buy and sell slots.[2] Questions remain about the optimal way to allocate existing or new slots. The first question asks why the incumbents should be given valuable property rights.[3] The second asks whether allowing carriers to buy and sell slots will lead to monopolization of airports, with resulting higher fares. Repossessing slots and then selling them back to carriers would raise revenue and would ensure that those carriers willing to pay the most would acquire the slots. This has positive welfare implications in a perfectly competitive environment, but the impact is less clear if there is the potential for wielding market power.

This *Economic Review* analyzes competition for scarce inputs such as airport takeoff and landing slots. It describes the outcome of an auction of slots between two carriers, who may

■ 1 Only one major airport has been built in the United States since 1974, that being the new Denver facility.

■ 2 For further discussion, see Grether, Isaac, and Plott (1989).

■ 3 A slot at O'Hare recently rented for $66,000 per month. See "United Wins TWA Lease," *New York Times*, March 20, 1992.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

have existing slots. This allows us to evaluate the welfare implications of selling slots. It likewise describes the outcome of a merger or takeover battle in which two large incumbents seek to buy up the slots of small competitors, if not the competitors themselves.[4]

The objects for sale (the slots) are sold in a sequence of auctions. The aggregate value of the slots, and each bidder's valuation of the marginal slot, increase as slots become more concentrated in the hands of one bidder. These conditions arise naturally in market games, since control of slots can confer market power by limiting the competitiveness of one's rivals. I focus primarily on the case of two bidders who start with no slots. The final allocation is typically unequal, and many different intermediate allocations lead to the same final allocation. In addition, if bidders start with nonzero holdings of slots, a bidder who holds the entire initial allocation of slots often does not block new entry. I then discuss the extension to more than two bidders, starting with three bidders, and then with more bidders than slots.

While the focus here is on the allocation of scarce inputs, the analysis contributes to the general theory of auctions as well.[5] Most theoretical studies of auctions have considered "one-shot" auctions in which a single object is sold or several objects are sold simultaneously. In reality, related objects are often sold sequentially, either because sellers enter the market one at a time, or because it is practical for an individual seller to sell related objects as an ongoing process. (Goods auctioned sequentially include art, wine, procurement contracts, timber rights, and mineral rights.) McAfee and McMillan (1987) note that all levels of governments of western nations procure goods and services equal to 10 percent of gross national product annually, much of it by competitive bidding. Given the economic importance of these markets, further study is warranted.

A common feature of many of these settings is that the value of the object for sale, be it a contract to produce or a scarce input, depends on the other objects obtained. If the objects are identical, then this observation implies that the total value of the objects obtained is a nonlinear function of the quantity obtained, which is the case here.

■ 4 For an example, see "American Deal Will End Britt O'Hare Operation," *Chicago Tribune,* December 20, 1988.

■ 5 See McAfee and McMillan (1987) for a survey of the literature on sequential auctions.

I. The Model

Suppose that N identical objects called slots are offered for sale by one or more sellers.[6] Assume that $N \geq 2$ and N is an even number. The sellers could be a government agency wishing to allocate some or all of the takeoff and landing slots at an airport, or they could be carriers with small holdings of slots. (Large carriers may wish to take over the small carriers outright or they may simply wish to purchase their slots.) Two bidders compete for the slots.

I denote a typical allocation of slots by the ordered pair (x, y), where x denotes the number of slots currently held by bidder X, y denotes the number of slots currently held by bidder Y, and $0 \leq x + y \leq N$. If $x + y = N$, then (x, y) is a final allocation. At each nonfinal allocation, an auction allocates the next slot. After all slots are allocated, each bidder receives a "final payoff" that depends on the final allocation. This payoff can be thought of as the profit from production.

The specific auction format used is the second-price sealed-bid auction. In this auction, the high bid wins, and the winner pays the amount bid by the losing bidder. This format is used for ease of exposition because its outcome mimics that of the standard oral ascending-bid auction. Suppose, for example, that bidder X was willing to pay up to $100 for a slot, whereas bidder Y was willing to pay up to $80. In an oral ascending-bid auction, where the auctioneer raises the price, bidder Y should not drop out until the price hits $80 exactly. Thus, bidder X will win the auction and pay a price of (approximately) $80. If, instead, the bidders were asked to write down how much they were willing to bid, with the understanding that the bidder who submitted the higher figure would receive the slot at a price equal to the loser's figure, then the outcome should be the same. The latter scheme is the second-price sealed-bid auction.

A minimal requirement for an equilibrium of this game is that each bidder's strategy be optimal, given the strategy of the other bidder. This is the requirement that the strategies form a Nash equilibrium. A Nash equilibrium is self-enforcing in the sense that neither bidder has an incentive to change her strategy, given the strategy of the other bidder.

Unfortunately, there can be many Nash equilibria, so we must be more discerning.

■ 6 The model and the analysis borrow extensively from Gale and Stegeman (1993).

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

Suppose, for example, that bidder X values a slot at $10 while bidder Y values it at $8. In a second-price auction, it is a Nash equilibrium for X to bid $10 and for Y to bid any amount strictly below $10. Neither bidder has an incentive to change her bid. If bidder X submits a bid above Y's bid, then she receives a surplus equal to the difference between her valuation ($10) and Y's bid. This holds regardless of X's exact bid, as long as she wins. Bidder Y has no incentive to change since he loses and receives zero surplus currently, whereas if he were to outbid X, then he would lose $2, since he would pay $10 for a slot that he values at $8. Thus, any pair of bids ($10, b_Y) with $b_Y <$ $10 constitutes a Nash equilibrium, as does any pair of bids (b_X, $8) with $b_X >$ $8.

While there are many Nash equilibria, only one is "robust" in a certain sense. Once we consider the possibility that bidders might make a mistake and submit the wrong bid, then bidder X should bid $10 and bidder Y should bid $8. If, for example, Y bids $9.50, expecting that X will bid $10, but X makes a mistake and bids only $9, then Y wins and pays $9, yielding a net surplus of –$1. Likewise, if X bids $9, thinking that Y will bid $8, but Y makes a mistake and bids $9.50, then bidder X has lost $.50 in surplus that would have accrued had she bid $10. The bids b_X = $10 and b_Y = $8 are referred to as "dominant strategies," since they are optimal regardless of the other bidder's bid. I restrict attention to dominant strategies, which leaves each bidder with a unique bidding strategy.[7]

I study the equilibrium sequences of winners and the associated prices. (Tie bids produce multiple sequences of winners.) The bidders know the initial allocation of slots, the auction rules, and the final payoffs (the profits from production, given the final allocation of slots). I focus on the case in which both bidders start the auction with no slots, but consider other initial allocations as the analysis progresses.

A numerical example illustrates the workings of the sequential auction. Suppose that there are two slots for sale. The two firms receive a profit from production of $2 apiece if each of them owns one slot. A firm receives a profit of $5 if it controls both slots, in which case the other bidder receives zero. The slots are sold in a sequence of two second-price auctions. Suppose that bidder X wins the first auction. She will be willing to pay up to $3 for the second slot, which is her marginal valuation. Bidder Y is willing to pay up to $2, which is his marginal valuation. Bidder X will win the second auction and pay a price of $2. Given this behavior in the second auction, both bidders are willing to bid exactly $3 in the first auction. Therefore, the price of the first slot is $3 and the price of the second is $2. One bidder wins both, but both bidders receive net profits of zero.

Now suppose that owning one slot is worth $3 instead, all else equal. After the first auction, the winner has a marginal valuation of $2, whereas the loser has a marginal valuation of $3. The loser of the first auction will win the second and pay a price of $2 for that slot. Thus, neither bidder will be willing to pay more than $2 to win the first slot, since by losing he or she will win the second slot. Similarly, each bidder will be willing to pay up to $2 for the first slot. Since winning the first auction at any price strictly less than $2 is better than winning the second auction at a price of $2, it follows that both bidders will bid $2 for the first slot. This process generalizes to N slots. The equilibrium bids are determined by backwards recursion, as outlined below.

At any point during the sequence of auctions, the current allocation of slots is an ordered pair (x, y), where bidder X owns x slots and bidder Y owns y slots. The allocation (x, y) is followed by $(x + 1, y)$ or $(x, y + 1)$, depending on who wins at (x, y). For each number of slots, $x = 0, 1, ..., N$ and each bidder $i = X, Y$; $f_i(x)$ denotes the final payoff given to bidder i if the final allocation is $(x, N - x)$. Since the slots are assets, f_X is increasing in x and f_Y is decreasing. Each bidder i seeks to maximize the difference between her final payoff $f_i(x)$ and her total expenditure for slots purchased in auctions.

Let $V_i(x, y)$, $i = X, Y$, denote bidder i's valuation of allocation (x, y). In other words, $V_i(x, y)$ is bidder i's equilibrium payoff in the game commencing at (x, y). The valuation of a final allocation is the final payoff:

$$(1) \qquad V_i(x, N - x) \equiv f_i(x), \quad x = 0, 1, 2, \dots, N; \; i = X, Y.$$

The valuation of an intermediate allocation (x, y) depends on the valuations of its sequels $(x + 1, y)$ and $(x, y + 1)$. There is a unique valuation of (x, y) if each bidder adopts the strategy in auction (x, y) and in all subsequent auctions of bidding the increment to value resulting from obtaining the next slot. (As noted

7 Strictly speaking, I eliminate, iteratively, all strategies that are dominated by some other strategy.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

above, any auction in which the resulting bids are unequal has many Nash equilibria, but the possibility of mistakes justifies the assumption that bidders will play dominant strategies.)

Consider first the penultimate allocations (x, y), where $0 \le x$, $0 \le y$, and $x + y = N - 1$. The value of that next slot to bidder X, say, is the difference between her final payoff when she wins and when she loses: $V_X(x + 1, y) - V_X(x, y + 1)$. It is a dominant strategy for bidder i to bid $B_i(x, y)$ in auction (x, y), where

$$\text{(2a)} \quad B_X(x, y) \equiv V_X(x+1, y) - V_X(x, y+1);$$

$$\text{(2b)} \quad B_Y(x, y) \equiv V_Y(x, y+1) - V_Y(x+1, y).$$

If $B_X(x, y) > B_Y(x, y)$, then bidder X wins the next slot and pays bidder Y's bid. If $B_X(x, y) < B_Y(x, y)$, then bidder Y wins the next slot. If $B_X(x, y) = B_Y(x, y)$, then the tie-breaking rule determines the winner.

The valuations at each penultimate allocation can now be calculated, given these equilibrium bids. This allows bids to be determined one auction earlier, at the antepenultimate allocations, and so on. Bids at any earlier allocation also take the form of (2). The bidders' valuations of each allocation can be calculated recursively. The price paid in auction (x, y) is min $[B_X(x, y), B_Y(x, y)]$. The tie-breaking rule, which I leave unspecified for now, cannot affect bids or valuations in any given auction.

The equilibrium of the auction game comprises the bidding functions $B_X(x, y)$ and $B_Y(x, y)$, for all (x, y). Henceforth, I assume that the bidders are symmetric with respect to their final payoffs. The outcomes will typically not be symmetric, however.

II. Competition for Scarce Inputs

I now put some additional structure on the model. First, symmetry of bidders is imposed in (3a). Second, aggregate final payoffs, and each bidder's valuation of the marginal slot, are assumed to rise as the final concentration of slots increases (that is, as the allocation of slots between bidders becomes more unequal). These conditions arise naturally for scarce inputs because higher concentration typically leads to higher output prices, increasing both aggregate profits and the value of marginal output. Assumptions (3b) and (3c) mean that each bidder's valuation of the marginal slot rises as the final concentration of slots increases, while assumption (3d) means that aggregate final payoffs increase with the final concentration of slots. (Note that [3c] says that bidder Y's final payoff is increasing and concave in his own number of slots.) Finally, assumptions (3e) and (3f) are invoked for concreteness. The example discussed below satisfies these conditions, but the results of Theorem 1 do not require them.

(3a) $f_X(x) = f_Y(N - x)$ for all x.

(3b) $f_X(x)$ is strictly increasing and convex, for $x \ge N/2$.

(3c) $f_Y(x)$ is strictly decreasing and concave, for $x \ge N/2$.

(3d) $f(x) \equiv f_X(x) + f_Y(x)$ is strictly increasing, for $x \ge N/2$.

(3e) $f_X(x) = \alpha_X + \beta_X x + \delta_X x^2/2$, and

(3f) $f_Y(x) = \alpha_Y + \beta_Y x - \delta_Y x^2/2$,

where $\delta_Y > \delta_X > 0$.

Assumptions (3a) – (3d) place additional restrictions on the parameters, which I leave unspecified for now.

Winning a particular slot affects a bidder directly by increasing her current holdings of slots, but it also affects the competition for the remaining slots. Two countervailing effects determine the number of slots won by, say, bidder X. The first effect argues for one bidder to acquire all slots, since that outcome maximizes the sum of the final payoffs. This effect is countered by the fact that the more slots bidder X wins, the higher is the marginal valuation of another slot to bidder Y (presuming that $y < x$), and the fiercer is the competition for the marginal slot. If bidder Y ends up with very few slots, then bidder X is a near-monopolist. Bidder Y benefits from X's relative passivity in the production game, so marginal capacity is most valuable to bidder Y when he has very little. (Empirically, airlines that dominate airports tend to have the highest fares. See Borenstein [1992].) This trade-off leads to an equilibrium allocation of slots that is typically neither symmetric nor monopolistic.

A heuristic argument provides some intuition for the determinants of the equilibrium allocation of slots. I proceed by conjecturing the form

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

of the equilibrium. I then show that it is optimal in each auction for bidders to bid in the way prescribed. The heuristic argument for determining the equilibrium allocation is now given.

Consider an allocation (x, y) such that, in equilibrium, bidder X wins any auction $(x', y') \geq (x, y)$ such that $y' \leq y + 2$. In other words, if bidder X has at least x slots and bidder Y has y, $y + 1$, or $y + 2$ slots, then bidder X will win all remaining slots in equilibrium. If a deviation at (x, y) causes bidder Y to win one more slot, then any additional slot that bidder Y wins will, in equilibrium, be his last slot. Therefore, following the deviation, bidder Y bids the value of that last slot, $B_Y(N - y - 2, y + 1)$, in every remaining auction. If the deviation did not occur, bidder Y would bid $B_Y(N - y - 1, y)$ in every remaining auction. Since bidder X wins all remaining slots, even if a deviation causes her to lose one of those slots, the marginal contribution to her final payoff of winning auction (x, y) is equal to the marginal value of winning the last auction, $B_X(N - y - 1, y)$.

By winning auction (x, y), bidder X increases the price that she must pay upon winning the remaining $N - x - y - 1$ auctions by $\Delta B_Y(N - y - 2, y + 1) \equiv B_Y(N - y - 1, y) - B_Y(N - y - 2, y + 1)$. Therefore, bidder X wins auction (x, y) if and only if $B_X(N - y - 1, y) \geq (N - x - y - 1)\Delta B_Y(N - y - 2, y + 1) + B_Y(N - y - 1, y)$. Recalling that $f(x) = f_X(x) + f_Y(x)$, the condition can be rewritten:

$$(4) \quad f(N - y) - f(N - y - 1) \geq (N - x - y - 1)\,\Delta B_Y(N - y - 2, y + 1).$$

In other words, bidder X wins auction (x, y) only if the increment to aggregate final payoffs (the left-hand side) exceeds the increment to total prices paid (the right-hand side).

Define μ to be the smallest integer such that $(y + 2, y)$ satisfies (4) for $y = \mu, \mu + 1, N/2 - 2$. The equilibrium allocation will be $(N - \mu, \mu)$, which I call the "modal allocation," or its symmetric counterpart.

Theorem 1. When (3) holds, either bidder X wins $N - \mu$ slots and bidder Y wins μ, or the reverse holds.

The equilibria of the original game and the game starting from any intermediate allocation have a regular form. From here on, I assume that ties are won by bidder X if the allocation is even. This tie-breaking rule ensures that bidder X always has at least as many slots as bidder Y, so the equilibrium allocation in this case is $(N - \mu, \mu)$. If $x = y$, then there will be a tie in the next auction, which buyer X wins by the tie-breaking rule. Buyer X also wins the next auction to maintain her lead. Thereafter, different patterns are possible. If I assume that bidder Y wins all ties when he is behind, then the bidders alternate victories until buyer X commences her final string of victories, and the modal allocation is reached. If $x > y$, then buyer Y wins a string of auctions first, but not enough to catch up. Either buyer Y catches up to within one slot, and then a pattern of alternation persists until the modal allocation is reached, or else he never catches up to within one, in which case buyer X wins all auctions after Y's string of victories. The point at which buyer X commences her string of victories is determined by (4).

The theorem is illustrated by the following numerical example. Suppose that each slot represents one unit of capacity in a subsequent production game, where the inverse demand is $P = 24 - Q$ and average cost is zero, up to capacity. If bidder X has x slots, then she is able to produce up to x units of output. For quantities strictly less than x, marginal and average cost both equal zero, but for quantities strictly above x, marginal and average cost are infinite.[8]

Suppose that the firms behave like Cournot duopolists in the production game. In a Cournot duopoly, each firm chooses the optimal quantity to produce, given the level produced by its competitor. In the absence of capacity constraints, each duopolist produces eight units of output in the Cournot equilibrium. To see this, note that if bidder X produces q_X units of output, then bidder Y's profit from producing q units is $(24 - q_X - q)\,q$, a concave function that takes its maximum at $q = (24 - q_X)/2$. Likewise, if Y produces q_Y, bidder X's optimal strategy is to produce $q = (24 - q_Y)/2$. If both bidders are unconstrained, then the unique Cournot equilibrium has $q_X = q_Y = 8$.

Assume, however, that exactly $N = 16$ slots are available, and $x > 8$. Concavity of the profit function ensures that the best responses for the two bidders are

$$(5a) \quad q_X = \min\{(24 - q_Y)/2, x\}$$

$$(5b) \quad q_Y = \min\{(24 - q_X)/2, 16 - x\}.$$

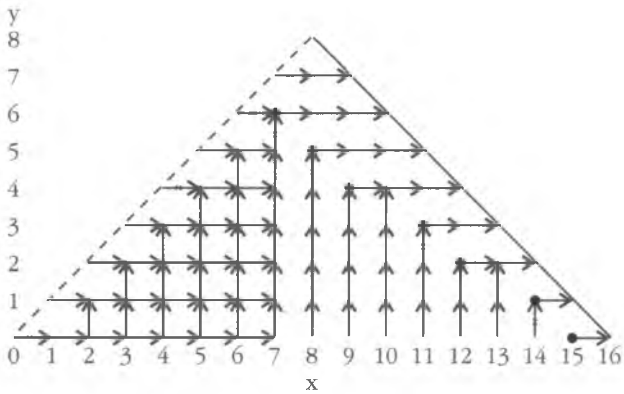
There are now four possibilities, depending on whether the bidders are constrained. The bidders cannot both be unconstrained, since that would require that bidder Y produce eight units, which exceeds his capacity. It is also immediate

8 In reality, the capacity constraint is not absolute, since more seats can be put on an airplane, but the qualitative properties need not change when this possibility is permitted.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

FIGURE 1

Possible Sequences of Equilibrium Victories



SOURCE: Author's calculations.

that the bidders cannot both be constrained. If Y produced $16 - x$ units, X would wish to produce only $[24 - (16 - x)]/2 = 4 + (x/2)$, whereas he has available x units of capacity. Because $x > 8$, $4 + (x/2) < x$. Now suppose that X is constrained and Y is unconstrained. If $q_X = x$, $q_Y = \min\{(24 - x)/2, 16 - x\} = 16 - x$, which is a contradiction. The only remaining possibility is that bidder X is unconstrained and bidder Y is constrained. It is straightforward to confirm that $q_X = 4 + (x/2)$ and $q_Y = 16 - x$ are the unique equilibrium outputs. Total output is $20 - (x/2)$. The price of output is $(8 + x)/2$, so the final payoffs (that is, the profits in the Cournot production game) are

(6a) $f_X(x) = (8 + x)^2/4$,

(6b) $f_Y(x) = (8 + x)(16 - x)/2$.

Note that these final payoffs are consistent with (3a) – (3f).

I now return to the auction. The different possible sequences of equilibrium victories are noted in figure 1. Arrows indicate which allocations can follow a given allocation. For example, starting from the allocation (2,0), where bidder X currently has two slots and bidder Y currently has none, the allocations (3,0) and (2,1) can both be reached. This means that the two bidders submit the same bid at (2,0), and the tie-breaking rule determines whether X wins and (3,0) is reached or whether Y wins and (2,1) is reached.

The modal allocation is (10,6) here. Straightforward calculations confirm that (4) amounts to

(7) $(2y + 1)/4 \geq (13 - 2y)$,

or $y \geq 5.1$. Thus, $\mu = 6$. Price drops occur between auctions separated by double cross marks. On any path from the origin, the price is 10.5 in the first auction, 6.25 in the next 12 auctions, and 5.5 in the last three auctions. The equilibrium paths in the example have several other properties.

1. While the leader may ultimately win considerably more slots than the follower, the leader strictly outbids the follower only when his lead is reduced to one or after the follower has already won all of her slots. A consequence is that there is an equilibrium path in which bidder X wins the first two slots, then bidder Y wins one, and then they alternate victories until the allocation $(\mu + 1, \mu)$ is reached. At that point, bidder X wins all remaining slots.

2. The trade-off that determines the number of slots that the leader wants to win, along with the preceding observation, ensures that many intermediate allocations lead to the same final allocation. Any initial allocation (x, y) at which $0 \leq y \leq x \leq \mu + 1$ and $(x, y) \neq (\mu + 1, \mu + 1)$ leads to the same final allocation. In the example, the modal allocation is (10,6), and any initial allocation (x, y) with $0 \leq y \leq 7$, excepting only (7, 7), leads to the modal allocation. Thus, many different histories lead to the same equilibrium allocation of slots. This implies that a sizable initial advantage need not be maintained.

3. If the initial allocation is (0,0), then the price drops either once or twice. It drops immediately after the leader is established (which equalizes net profits). It also declines immediately after the follower wins his last slot if a deviation causing him to lose that slot would have caused him to lose all subsequent slots as well. Since the follower loses some surplus in that case, he bids more than he would bid if a loss would be compensated by his winning another auction subsequently. The existence of at most three distinct prices depends on the precise functional form assumed.

4. An incumbent monopolist typically does not block entry if new slots become available. For instance, suppose that the initial allocation of slots is (12,0), which effectively makes firm X an unconstrained monopolist. Calculations show that if four new slots become available, then the prospective entrant, firm Y, wins the first two slots before firm X wins the last two.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

The reason for this is simply that it is often too costly to preempt an entrant completely, since the incumbent monopolist would leave the capacity idle whereas the entrant would find the capacity very valuable because the incumbent would still be a near-monopolist after entry.

III. More Bidders

In many markets, there are exactly two competitors of consequence. This is particularly true at many airports and in many airline markets, so the assumption of two bidders is realistic in such contexts. I now consider the impact of having more than two bidders. The extension is qualitatively different because the value of winning a slot typically depends on which rival would otherwise win it.[9] Hence, a bidder often has no dominant strategy, so I focus on the equilibria in which each bidder bids the value of the next slot to her, assuming that the rival making the highest bid would otherwise win the auction.

Multiple equilibria are common when productive capacity is sold, because keeping capacity away from constrained bidders is a public good that the unconstrained bidders may individually prefer not to provide. Thus, at an airport with two dominant incumbents, if new capacity becomes available, each incumbent may prefer that the other buy the new capacity and keep the entrant out. The following example illustrates the public good aspect in a case where all bidders start with no capacity.

Suppose that three firms bid for three slots. Each slot represents one unit of capacity, where inverse demand is $P = 8\,(4 - Q)$ and production is again costless up to the capacity constraint. In the post-auction production game, the firms again act as Cournot competitors. The final allocation (3,0,0) returns final payoffs of 32, zero, and zero, respectively. The allocation (2,1,0) returns payments of 18, 12, and zero. The allocation (1,1,1) returns a payment of 8 to each bidder. The payments from the other final allocations are determined symmetrically.

Equilibria are computed through backwards recursion. (The price paid in the second-price sealed-bid auction is the larger of the two losing bids here.) Consider the allocation (2,0,0). The last slot is worth 14 to the leader and bidder X, and 12 to the others. Therefore, the leader pays 12 for the last slot. The payoff to entering this "subgame" (the game that takes place when the initial allocation is (2,0,0)) is 32 – 12 = 20 for the leader and zero for the followers.

Consider next the allocation (1,1,0). The last slot is worth 10 to a leader if it would otherwise go to the follower, bidder Z, but it is worth only 6 if it would otherwise go to the other leader. It is worth 8 to the follower. If tie bids are broken fairly, then there are three equilibria of interest. The follower bids 8 in all three equilibria. In one equilibrium, one leader bids 10 and the other 6; in another, their roles are reversed; in the last, they both bid 8. In the first two equilibria, the payoff to entering this subgame is 10 for the leader who wins, 12 for the other leader, and zero for the follower. Both leaders are willing to bid enough to shut out the follower, but each prefers that the other do it. If we assume that the leaders randomly coordinate on one of the two equilibria, the average payoff to entering this subgame is thus 11 for each leader and zero for the follower.

At the allocation (1,0,0), the next slot is worth (32 – 12) – 11 = 9 to the leader and 11 to each follower. (If the leader wins, he goes on to win the third slot at a price of 12.) One of the followers wins, and the payoffs to entering this subgame are 11 for the leader and zero for the followers. At the initial allocation (0,0,0), each firm bids 11.

Summarizing, the assumption that firms randomly coordinate among equilibria at the time of the auction generates an equilibrium allocation that is a permutation of (2,1,0). The final payoffs to the firms are 18, 12, and zero, and the equilibrium prices are $p_1 = 11$, $p_2 = 11$, and $p_3 = 8$.

It is possible for three firms to earn higher profits if they coordinate on an equilibrium before the auctions begin. In particular, suppose that they coordinate on the equilibrium in which firm X wins auctions (1,1,0) and (1,0,1) and firm Y wins auction (0,1,1). Then the leader's payoffs in these subgames are either 10 or 12, instead of 11. In auction (1,0,0) the equilibrium price is 12, and in auctions (0,1,0) and (0,0,1) the equilibrium price is 10. The payoffs in these three subgames are now 10 for the leader and zero for the followers, except that firm Z's payoff in subgame (0,1,0) is 2 instead of zero, and its payoff in subgame (0,0,1) is 12 instead of 10. If tie bids are broken fairly, then firm Z wins the initial auction (0,0,0) with a bid of 11 and pays a price of 10. Since the payoff to winning is 12, firm Z earns profits of 2, unlike the zero profits earned without advance

9 This is reminiscent of the situation in baseball pennant races, where contending teams attempt to trade for a player whose team is out of the race. How much a team is willing to pay depends on whether the other team vying for that player's services is in the same division.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

coordination. The price sequence is $p_1 = 10$, $p_2 = 10$, $p_3 = 8$. Profits appear because advance coordination creates an asymmetry that gives firm Z an advantage from the start. This holds because firm Z is never put in the position of purchasing capacity that will not be fully utilized, since the other leader is assumed to win at (1,0,1) and at (0,1,1). In those cases, the other leader produces only 1.5 units even though it purchases 2.

Finally, consider the possibility of there being more bidders than slots available. In Krishna (1993), an incumbent monopolist and potential entrants bid for new capacity. The monopolist wins only the last unit of new capacity, assuming that marginal costs are constant, market demand is concave, and entrants always produce to capacity. The presence of potential entrants, each of whom will produce up to capacity, means that the bid from an entrant is always the equilibrium price of output should the entrant win. With two or three bidders, all bidders may have some slots, in which case they all internalize the impact of increased (aggregate) production on the value of their current holdings. This makes the potential entrants less competitive in the small numbers case, so an incumbent monopolist will win more slots.

IV. Concluding Remarks

This paper has provided a framework for analyzing competition for scarce inputs such as airport takeoff and landing slots. The analysis describes the outcome of an auction of slots between two carriers, who may have existing slots, and it also depicts the outcome of a merger or takeover wave. The equilibrium allocation of slots is typically asymmetric, even though firms are ex ante identical. It is not typically monopolistic, however, since the more concentrated the allocation of slots becomes, the higher is the price that the leader must pay for the marginal slot.

Many different histories, and many different allocations of slots, lead to the same equilibrium allocation of slots, implying that an initial advantage need not be maintained. Thus, the concern with monopolization may be misplaced. Future work will consider risk aversion, capital constraints, and matching of slots between airports.

References

Borenstein, Severin. "The Evolution of U.S. Airline Competition," *Journal of Economic Perspectives*, vol. 6, no. 2 (Spring 1992), pp. 45–73.

Gale, Ian, and Mark Stegeman. "Sequential Auctions of Endogenously Valued Objects," University of Wisconsin, Social Systems Research Institute working paper, 1993.

Grether, David M., R. Mark Isaac, and Charles R. Plott. *The Allocation of Scarce Resources: Experimental Economics and the Problem of Allocating Airport Slots*. Boulder, Colo.: Westview Press, 1989.

Krishna, Kala. "Auctions with Endogenous Valuations: The Persistence of Monopoly Revisited," *American Economic Review*, vol. 83, no. 1 (March 1993), pp. 147–160.

McAfee, R. Preston, and John McMillan. "Auctions and Bidding," *Journal of Economic Literature*, vol. 25, no. 2 (June 1987), pp. 699–738.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

Regional Wage Convergence and Divergence: Adjusting Wages for Cost-of-Living Differences

by Randall W. Eberts and Mark E. Schweitzer

Randall W. Eberts is the director of the W.E. Upjohn Institute for Employment Research in Kalamazoo, Michigan, and Mark E. Schweitzer is an economist at the Federal Reserve Bank of Cleveland.

Introduction

One of the basic tenets of economics is that the mobility of labor and capital tends to equalize prices across markets. This tendency toward price convergence is particularly notable across regional markets in the United States. For as long as regional income data have been collected, per capita income and wage rates have generally become more alike.[1]

In light of this long-run trend, a surprising reversal has occurred in several regional price measures. Since the early 1980s, the regional dispersion of wages, housing prices, and the general cost-of-living indexes has been on the rise. Browne (1989) provides evidence that regional disparities in per capita income have been widening, while Eberts (1989) finds an increase in regional wage dispersion. In addition, we demonstrate below that housing costs and regional price indexes have been following a similar pattern. Curiously, however, wages adjusted for regional cost-of-living differences (which for brevity we refer to as *locally adjusted* wages) have continued to converge.

Temporary deviations from the tendency toward convergence are not unexpected, as localized shocks can result in significant adjustments to regional prices. Eberts and Stone (1992) and Blanchard and Katz (1992) show that negative localized employment shocks to a metropolitan area can depress wages there by as much as 40 percent of their original level for up to six years before equilibrium returns. Even so, a significant period of increasing dispersion, as observed in the 1980s, is rare. In the last century, regional per capita income diverged only one other time, between 1920 and 1940.

This paper focuses on the details of regional convergence or divergence in goods prices, nationally adjusted wages (wages deflated according to the national price level), and locally adjusted wages. Our goal is to identify and describe these obviously related phenomena. The characterization of this relationship follows Roback's (1982) model of equilibrium in local labor and land markets in the presence of local quality-of-life and production differences.

The dispersion in locally adjusted wages depends on the dispersion of its components: nationally adjusted wages and local prices. We

1 Eberts (1989) demonstrates this trend. Unfortunately, the regional wage series is relatively short, beginning only in the 1950s. However, the same general pattern is found in regional per capita income, which is largely composed of wages and which extends well into the 1800s.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

demonstrate the linkage between wages and prices by showing how the comovements of nationally adjusted wages and regional prices affect locally adjusted wages. Given that the two components of locally adjusted wage variation have followed similar paths, it is the growing covariance of these measures that results in continued convergence of locally adjusted wages between census regions.

We also show that trends in the two wage dispersion series primarily reflect regional differences in market valuations of worker characteristics rather than shifts in the levels of workforce characteristics. We modify the decomposition used by Eberts (1989) in examining the U-turn in nominal wage dispersion. He identifies two factors: 1) regional differences in the return on various worker attributes and in wage differentials among industries and occupations, and 2) regional differences in the level of worker attributes and in the distribution of workers among industries and occupations. Basically, these two factors distinguish between wage dispersion caused by regional markets placing different values on identical attributes, and dispersion caused by regions having different compositions of attributes, even though regional markets value these attributes similarly.

The analysis supports previous studies showing that changes in regional wage differentials over time result from varying valuations of worker attributes, not from shifts in the regional composition of the workforce. The additional insight offered by this paper is that market forces produce different patterns of regional dispersion of nationally and locally adjusted wages. While not directly explained here, these differences are consistent with the view that workers and businesses pursue separate objectives or place unequal weights on local prices and amenities.

I. Explaining Regional Wage and Price Differentials

The key to understanding potentially permanent regional wage differentials is to recognize that not all factors are mobile across regions. Workers and firms interact in regional labor markets, determining wages and prices. Although firms and their employees may respond quickly to changes in local market conditions, some factors that are unique to a region, such as geographic and climatic characteristics, remain the same. Even for those areas that share common features, the quality and quantity of site-specific characteristics may differ. Therefore, firms and households may be willing to pay or accept different levels of compensation depending on the value they place on those attributes. These immobile, site-specific features are referred to here as amenities: consumptive amenities apply to households and productive amenities apply to firms.

A few examples of potential sources of consumptive and productive amenities indicate their conceptual breadth and complexity. The prototypical consumptive amenity is a weather advantage. California and Florida attract people who prefer a warm climate and who are willing to accept the higher costs of living there. Other potential consumptive amenities include familial or historical ties to an area, region-specific recreational activities (skiing or surfing, for example), community spirit, and the quality and age of the housing stock. Despite the positive connotation of the term amenities, in our usage it also encompasses the negative features of an area, such as high crime rates or a combination of high local taxes and poor local government services.[2]

Port facilities are an excellent example of a productive amenity, since they can lower transportation costs for firms located nearby. Productive amenities also include low-cost distribution channels, informational advantages provided by firms' proximity to other similar producers or suppliers, and state or local government protections or restrictions pertaining to local businesses.

Interpreting Wage and Price Dispersion

Interpreting regional wage and price convergence in this framework is difficult. Households and businesses can and will move to locations where they can better prosper. If both labor and capital are mobile, factor prices could converge or diverge in response to shifts in either firms' and workers' valuation of local amenities or changes in the availability of amenities in various locales.

Another source of apparent convergence or divergence in regional wages and prices is the economy's constant adjustment to a stream of shocks. The demand for and supply of labor in

■ 2 Local taxes are potential negative amenities to the extent that they are not included in prices. However, property taxes are essentially incorporated into the Consumer Price Index under the "rental equivalence measure" of housing costs.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

an area may be radically altered by technological changes or shifts in consumption preferences. Although households and businesses are mobile, adjustment delays may result in temporary periods of divergence. Studies by Eberts and Stone (1992) and Blanchard and Katz (1992) suggest that the adjustment period to a local labor-market shock may be as long as 10 years.

Since housing and locally produced goods and services represent a major portion of a household's budget, these prices become an important component of household utility and thus of household decisions. If local goods accounted for the entire household budget, then consumption would equal household wages deflated by local prices. On this basis, we assume that given a stable value for local amenities, locally adjusted wages represent the primary motivator of household mobility.

By contrast, the price of local goods and services, including land, plays a smaller role in business decisions. Wages are generally a larger fraction of most firms' costs than are local goods. Furthermore, for producers of local goods, an increase in local prices would affect both revenues and costs. The marginal firm most likely to relocate would be a producer of national goods for whom any rise in local wages (or prices) relative to those faced by its competitors would immediately lower profits. For these firms, local wages (nationally adjusted), with little regard for local prices, should be the determining factor in their location once amenities and previous capital investments have been accounted for.

It is important to compare observationally equivalent workers if we are to measure regional wage differentials accurately. Firms consider the skill level of their workforce as well as the size of their payrolls when making location decisions. Similarly, workers must evaluate the marketability of their skills in various regions when comparing locally adjusted wages. Therefore, regional shifts in factors associated with worker productivity, such as average educational attainment or workforce experience, should be controlled for in any analysis of factor-price adjustments. Shifting patterns of employment by industry or occupation, which may be related to compensating differentials associated with features of those jobs, should also be considered.

II. Wage and Price Trends

Wages

Wages of individual workers are obtained from the March *Current Population Survey's* (CPS) wage supplements for the years 1973 through 1991. The March survey reports annual wage and salary data and weeks worked from the previous year. Dividing annual earnings by weeks worked yields average weekly earnings for the years 1972 and 1990. For purposes of the respondents' confidentiality, these data are coded by the Bureau of Labor Statistics (BLS) with a maximum salary for individuals whose pay exceeds the top-code value (for example, $199,998 after 1989). Average weekly earnings are computed after correcting for top-coding by assigning these individuals the mean of the appropriate Pareto distribution.[3] The sample is limited to full-time workers who were employed all year or who, if unemployed for part of the year, spent that time actively seeking work. Because only full-time workers are included, average weekly wages closely approximate average hourly wages.

Two definitions of regions are used in this paper: Metropolitan Statistical Areas (MSAs) and census regions. Since cost-of-living indexes are available only for metropolitan areas, the basic unit of analysis is the MSA. The CPS identifies 44 MSAs, but the limited availability of price data for some of them reduces the usable number to 21. The small number of respondents in most MSAs lowers the efficiency of estimation for that section of the analysis dealing with the sources of wage convergence and divergence. To increase the number of individuals sampled in a given period, we pool together three years of individual responses for each MSA, resulting in a much broader coverage of worker characteristics and wages. Each of our six periods is identified by the middle year of the pooled three-year sample.[4] For example, the first period, which consists of earnings in 1972, 1973, and 1974, is referenced as 1973 in the figures and tables.

To provide another means of increasing the sample size for geographic comparisons, as

■ **3** See Shryock and Siegel (1971) for details on how a Pareto distribution may be applied to truncated wage data. The Pareto distribution assumes an exponential decline in the number of individuals with incomes above a certain amount, which is a reasonable characterization of higher income levels.

■ **4** The final period covers 1987 to 1990.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

FIGURE 1

Variance of Regional Log Wage Differentials: Single vs. Grouped Years

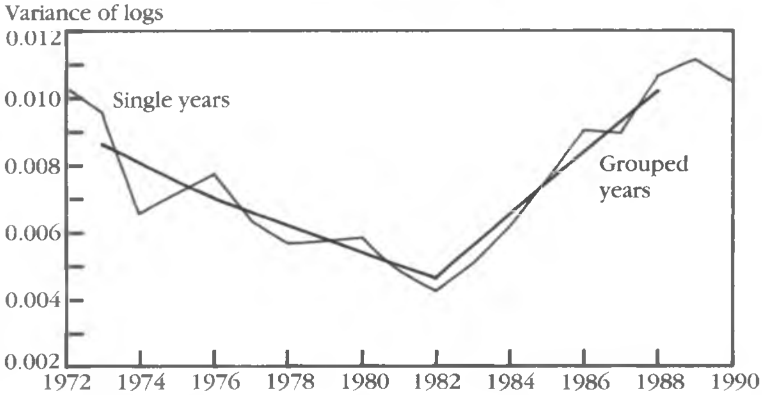


FIGURE 2

Variance of Regional Log Wage Differentials: Regions vs. Weighted MSAs

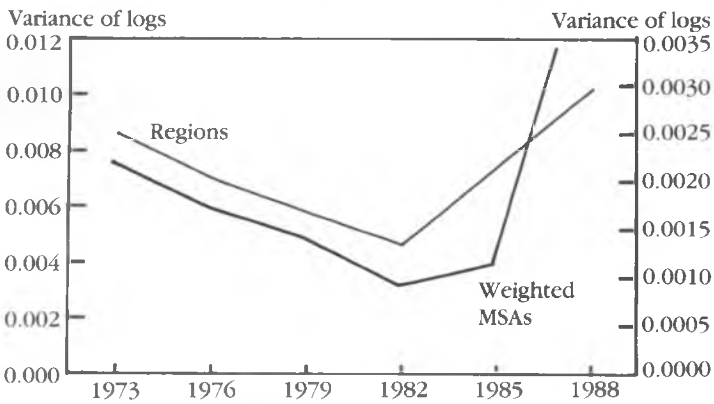
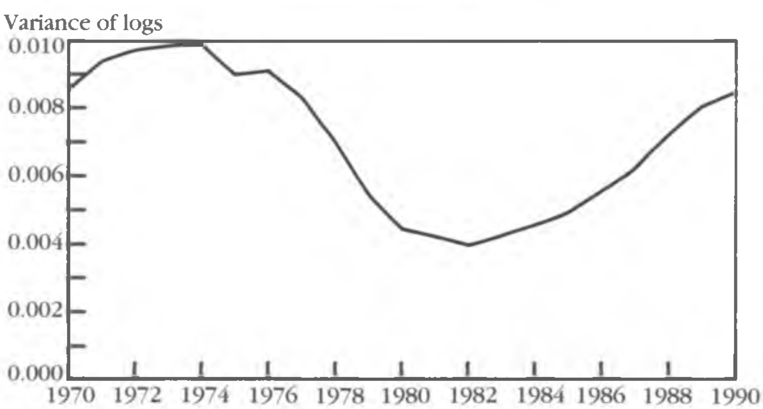


FIGURE 3

Variance of Local Price Indexes



SOURCE: Authors' calculations.

well as to be consistent with earlier work by Eberts (1989) and Browne (1989), MSAs are aggregated by proportional population weighting to represent the nine census regions. Each of these regions contains at least one of the 21 MSAs, except for the East South Central states (Kentucky, Tennessee, Alabama, and Mississippi). As shown below, the patterns of wage and price dispersion for MSAs and the constructed census regions are quite similar. To adjust for the effects of inflation, wages are deflated to 1982 levels by the GDP implicit price deflator.

Wage variance across regions exhibits a marked U-shaped pattern between 1972 and 1990, with wages converging during the first half of the period and then diverging thereafter (figure 1). From 1972 to the trough, the variance of wages is cut roughly in half. By the end of the period, the variance surpasses the level at which it started in the early 1970s. This convergence and subsequent divergence is apparent for single and grouped years.

The same basic pattern of wage dispersion is found in the MSAs aggregated to simulate the census regions (figure 2). The level is generally lower for the weighted MSA results because metropolitan wages are more alike across regions — even though major MSA wages generally reflect their region's differential. The differences between the two variances, shown in figure 2, reflect the degree to which regional wage differentials are altered by including smaller MSAs and rural areas. These patterns are generally consistent with the convergence/divergence phenomenon reported by Eberts (1989) using the May CPS, and by Browne (1989) using per capita income (of which wages account for a large portion).

Prices

Indexes that measure regional cost-of-living differences as well as price changes over an extended period are not readily available. The BLS releases a Consumer Price Index for selected MSAs that records price changes for each area over time. However, the index is constructed to ignore price differences across these MSAs by benchmarking the series to 100 on the basis of 1982 to 1984 prices within each area. In order to include this component in a regional price index, we rebenchmarked these indexes using the relative metropolitan cost-of-living index from the 1981 BLS Report on Family Budgets. (The report has not been updated because funding for the project was eliminated.) The

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

FIGURE 4

Locally Adjusted Wage Variation

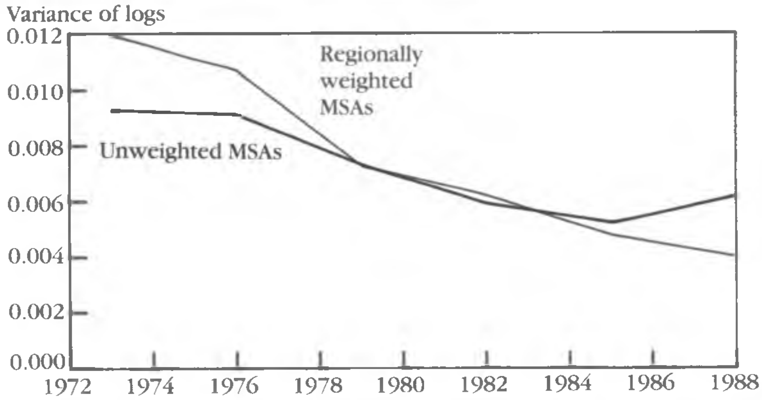
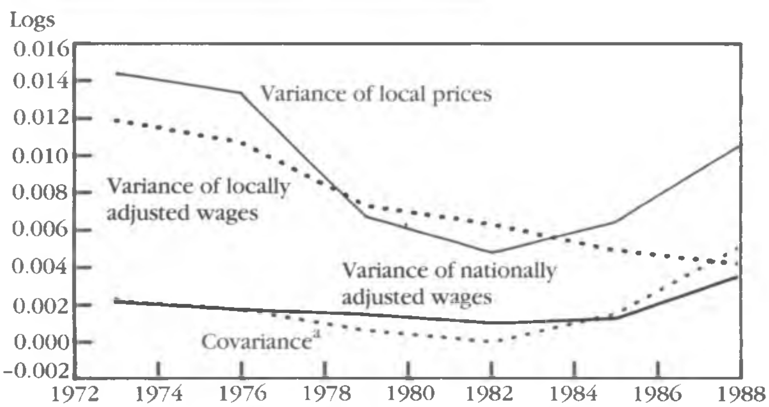


FIGURE 5

Factors in the Convergence of Locally Adjusted Wages



a. Covariance of nationally adjusted wages, local prices.
SOURCE: Authors' calculations.

metropolitan cost-of-living differences are based on a consumption basket appropriate for a four-person family with an intermediate income. The index that we construct identifies metropolitan price differences over time, which we use as our deflator instead of relying on a national price deflator.[5]

Figure 3 tracks the variance in the log of the metropolitan price index over the same period as wages. We use the log form to be consistent with the use of log wages to measure wage dispersion. Note that the dispersion of local prices follows a similar U-shaped path, declining during the first half of the period, reaching its nadir in the early 1980s, and then returning to previous levels. The dispersion of prices aggregated to the nine census regions exhibits a similar pattern.

The largest component of the regional price index — and the one that accounts for most of the difference in prices across MSAs — is the cost of housing services. This measure is closely linked to the general price index, as indicated by correlations between the MSAs' relative prices and housing costs of greater than 0.95 in each period. Thus, locally adjusted wages can be viewed as wages adjusted for local housing prices.

Locally Adjusted Wages

Locally adjusted wages refer to wages divided by local prices (including cost-of-living differences between localities). For comparisons between census regions, these wages are aggregated in the same fashion as regional wages and prices Locally adjusted wages do not conform to the marked pattern of regional convergence/divergence found in nationally adjusted wages and prices. Rather, the measure generally converges throughout the entire period. This is most pronounced for MSAs aggregated to the nine census regions, as shown in figure 4. From peak to trough, the variance of locally adjusted wages declines by almost 50 percent. This tendency toward convergence is confirmed at the metropolitan level for locally adjusted wages, except for a slight increase in the last period. In order to be consistent with the previous literature, we focus on census regions below.

III. The Relationship between Wages, Prices, and Locally Adjusted Wages

Figure 5 offers a complete picture of locally adjusted wages and its two components. As discussed earlier, the variance of the log of nationally adjusted wages is considerably smaller than both the variance of log prices and the variance of the log of locally adjusted wages. The covariance of the log of nationally adjusted wages and prices is positive, but smaller than the individual variances. This positive but weak covariance suggests that MSAs with above-average rents also pay above-average wages, which is consistent with Gabriel, Shack-Marquez, and Wascher's (1988) finding that higher rents are only weakly associated with higher wages. The relationship of the dispersion of wages, prices, and locally

5 As already noted, wages are deflated by the GDP implicit price deflator to eliminate the effects of inflation.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

adjusted wages can be seen by decomposing the variance of the log of locally adjusted wages.

$$(1) \quad var[\ln(w_r/p_r)] = var[\ln(w_r)] + var[\ln(p_r)] - 2\,cov[\ln(w_r), \ln(p_r)],$$

where r refers to the region, w_r is the average wage in region r, and p_r is the relative price level in region r. The variances are calculated independently for each year. Thus, the change in the variance of price-adjusted wages between two time periods (0 and 1) can be decomposed for each region as follows, dropping the redundant r subscripts:

$$(2) \quad var[\ln(w_1/p_1)] - var[\ln(w_0/p_0)]$$
$$= \{var[\ln(w_1)] - var[\ln(w_0)]\}$$
$$+ \{var[\ln(p_1)] - var[\ln(p_0)]\}$$
$$- 2\,\{cov[\ln(w_1), \ln(p_1)]$$
$$- cov[\ln(w_0), \ln(p_0)]\}.$$

Note that price-adjusted wages can converge, while nominal wages and prices diverge only if the difference in the covariances is at least half the magnitude of the two differences in variances. Local amenities, whether the benefits are accrued by firms or workers, provide a link between wages and prices. Thus, there are substantial grounds for the existence of non-negligible covariances between relative wages and prices. Figure 5 does identify a significant covariance, the rise of which is coincident with the increase in wage variation among regions.

IV. Does Regional Wage Convergence/ Divergence Represent Shifting Incentives?

In order to explore regional wage differences, observationally equivalent workers must be compared. The role of regional workforce differences in the relative wages of regions should be isolated from pay differentials that comparable workers would receive in other regions. We account for most sources of wage disparity by evaluating the typical differences in returns associated with worker characteristics, including education levels, experience, industry, race, and sex. The dispersion of regional wage differentials over time is decomposed into two components: changes in worker characteristics and changes in labor market implicit valuations of worker characteristics (as measured by regression coefficients). Because we are not the first to attempt to account for workforce differences, we start by reviewing the existing literature.

Previous Studies

Previous studies examining the relative size of the two components of wage differentials have focused primarily on explaining differences between the South and other regions of the United States. Sahling and Smith (1983), for example, compare the southern states with four other regions of the country: the Northeast, the West, the North Central states, and the New York metropolitan area. They estimate separate price-adjusted and nominal wage equations using a sample of residents from 29 of the largest MSAs in these five regions. Worker-attribute variables include measures of schooling, experience, race, occupation, sex, industry, job status, and union membership. Using two cross sections of data from the May 1973 and May 1978 CPS, the authors conclude that cost-of-living adjustments dramatically increased the wages of southern workers relative to their counterparts across the United States.

Farber and Newman (1987) extend Sahling and Smith's analysis to look explicitly at changes in characteristic prices over time. In addition to analyzing regional wage differentials in two separate years (1973 and 1979), they estimate the changes in differentials between the two years for various pairs of regions. Their results show that more than half of the predicted shifts in South/non-South wage ratios can be accounted for by changing relative returns to worker characteristics.

Using the same framework adopted in the current paper, Eberts (1989) examines the sources of nominal regional wage convergence and divergence on a full sample of workers from the May CPS. He finds that differences in the returns to worker characteristics account for both the convergence in regional wages from 1973 until 1982 and the divergence thereafter.

Other studies, using similar techniques but more detailed data, do not necessarily agree with the conclusion that characteristic prices explain regional wage differentials. Bellante (1979) and Gerking and Weirick (1983), for example, find that regional wage differences result primarily

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

from variations in the level of worker characteristics. These findings leave open the possibility that both characteristic prices and levels are likely sources of regional wage differentials.[6]

Defining Sources of Wage Differentials

Following the human capital specification of Hanoch (1967) and Mincer (1974), we specify the logarithm of individual wages — expressed in either nominal or price-adjusted terms — as a function of various worker attributes, including education level (entered as dummy variables for the completion of four levels of schooling, from high school to graduate studies), and potential experience (age, minus years of education, minus six, entered as a quadratic). Dummy variables indicating race, gender, occupation, and industry are also included as recognized factors in individual earnings. Time dummies are incorporated to account for aggregate fluctuations, including the business cycle, within each of the pooled three-year periods.[7]

We estimate hedonic wage equations separately for each period and for each of the 21 MSAs. Prior to the estimation, individual wages are deflated by either the national or local price index, as described previously. We weight regional wages and estimated wage components by their respective population shares in order to construct a regional measure. The East South Central region is excluded from the analysis because no metropolitan area price data were available for cities in these states. We then compare the regional wage estimates to national estimates based on the same regression and the sample of workers from all 21 MSAs.

The technique used to account for the two sources of wage differentials follows the approach of Oaxaca (1973), with modifications by Sahling and Smith (1983). The decomposition assumes that y, the logarithm of wages, can be appropriately described as a function of the worker and industry characteristics discussed earlier (X_i) and the hedonic labor market valuation of each characteristic (b):

$$(3) \qquad y = bX_i + u_i .$$

Estimating a well-specified earnings equation for each region accounts for the value associated with regional concentrations of particular workforce traits by identifying the average valuation of these traits in the region ($\hat{b}_{St}$ for region S at time t). Using y for $\ln(w)$, we can decompose the percentage difference in wages between the regions during one time period as follows:

$$(4) \qquad (\overline{y_{St}} - \overline{y_{Nt}}) = (\hat{b}_{St} - \hat{b}_{Nt})\,\overline{X_{Nt}} + (\overline{X_{St}} - \overline{X_{Nt}})\,\hat{b}_{Nt} + (\hat{b}_{St} - \hat{b}_{Nt})\,(\overline{X_{St}} - \overline{X_{Nt}}).$$

The first term on the right side accounts for the difference in labor market valuations of worker attributes between a region and the base. The second term denotes the difference in levels of worker and industry characteristics. The third term, a remainder, is generally assumed to be small and in fact proved trivial in our analysis. Below, we examine the relative contribution of the first two right-side components of equation (4) over time for both wage series.

Decomposing the Variance of Regional Wages

Having decomposed the regional wage differentials into separate factors, a number of variance decompositions are possible. The traditional decomposition focuses on the variance of the first two terms of equation (4), neglecting the third term (the interaction term):

$$(5) \qquad var(\overline{y_{St}} - \overline{y_{Nt}}) = var[(\hat{b}_{St} - \hat{b}_{Nt})\,\overline{X_{Nt}}] + var[(\overline{X_{St}} - \overline{X_{Nt}})\,\hat{b}_{Nt}] + 2\,cov[(\hat{b}_{St} - \hat{b}_{Nt})\,\overline{X_{Nt}}, (\overline{X_{St}} - \overline{X_{Nt}})\,\hat{b}_{Nt}] + \text{interaction term}.$$

This approach generally yields the correct interpretation of the sources, although it is incomplete due to exclusion of the interaction term. If the covariance is significant but is not reported, then the decomposition is even less complete. We report the results of this decomposition for purposes of comparison with the existing literature.

The focus of our paper, however, is on regional wage differences when local prices are factored in, so it is valuable to consider how

■ 6 Dickie and Gerking (1988) provide a comprehensive and insightful critique of the literature.

■ 7 If business cycle fluctuations alter general earning levels, then ignoring that variation would result in inconsistent estimates. Dummy variables account for the mean aggregate differences between the two periods.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

price-level corrections affect the variance decompositions. The adjustment for local prices is applied to individual earnings as follows:

(6) $y^*_{it} = y_{it} / p_{St}$,

where p_{St} is constant within the locality at time t and y^*_{it} and y_{it} are the vectors of individual-level wage observations for region S at time t. Given that each MSA's wage equation is estimated independently for every three-year block of time, we can expect that the transformation of the dependent variable will adjust the $\hat{b}_{St}$ estimates proportionally. This follows from the normal equations for annual estimates:

(7) $\hat{b}^*_{St} = (X' X)^{-1} X' \frac{Y}{p_{St}} = \frac{1}{p_{St}} \hat{b}_{St}$.

With pooling over three years of data, the adjustment would be a weighted average of the relevant p_{St}'s. Thus, in terms of the decomposition of regional wage differentials shown in equation (4), only the price term $[(\hat{b}_{St} - \hat{b}_{Nt}) X_{Nt}]$ and the interaction term reflect the adjustment of wages for the local cost of living.

Consequently, only the valuation component of the variance decomposition (equation [5]) would be altered, perhaps indicating that cost-of-living adjustments affect the variance of regional wages through the valuation of skills alone. These variance terms, however, are not a complete decomposition of the sources of regional wage differentials unless the covariance between the valuation and workforce characteristic terms is zero. The covariance term represents the correlations between regional concentrations of labor skills (or other characteristics) and differentials paid to those skills. A simple supply and demand model without perfectly elastic or inelastic demand or supply would predict non-zero correlations. The covariance would then be reduced to the extent that labor or firm mobility eliminated either regional skill concentrations or the wage differentials paid to specific skills. However, regional production and consumption amenities should ensure that this covariance is non-zero.

The problems with the commonly applied decomposition in this context suggest the need for an alternative decomposition that accounts for the covariance term in a meaningful way. A more complete decomposition that satisfies this requirement is provided by the covariances between the dependent variable and the additive factors.[8] To simplify the notation of the price and quantity equation (4), let $\tilde{y}$ be regional differentials, $\tilde{b}$ be the valuation term, $\tilde{x}$ be the workforce characteristics term, and $\tilde{i}$ be the interaction term. The interaction term, which we still expect to be small, is included so that the definition of the decomposition is complete. In place of equation (5), applying a covariance decomposition to the factors shown in equation (4) results in

(8) $var(\tilde{y}) = cov(\tilde{y},\tilde{b}) + cov(\tilde{y},\tilde{x}) + cov(\tilde{y},\tilde{i})$.

The three decomposition terms in equation (8) are easily interpreted as the effect of a factor on the dependent variable after covariances with all other factors have been accounted for. The first term represents the effect of labor market valuations, the second represents the effect of labor force differences, and the third is the effect of the small interaction term. Factors can be either positive or negative, depending on whether they augment or offset the sum of the other factors contributing to the variation. If the factors are fully independent, then the decomposition simplifies to the basic variance decomposition for the independent variables shown in equation (5), with a covariance equal to zero.

Splitting the parenthetical terms in equation (8) distinguishes the components of the three terms. Equation (9) shows that each term of this decomposition includes an expression for the relationship between valuations and workforce characteristics.

$$
\begin{aligned}
(9) \quad var(\tilde{y}) = {} & [var(\tilde{b}) + cov(\tilde{b},\tilde{x}) + cov(\tilde{b},\tilde{i})] \\
& + [var(\tilde{x}) + cov(\tilde{b},\tilde{x}) + cov(\tilde{x},\tilde{i})] \\
& + [var(\tilde{i}) + cov(\tilde{b},\tilde{i}) + cov(\tilde{x},\tilde{i})].
\end{aligned}
$$

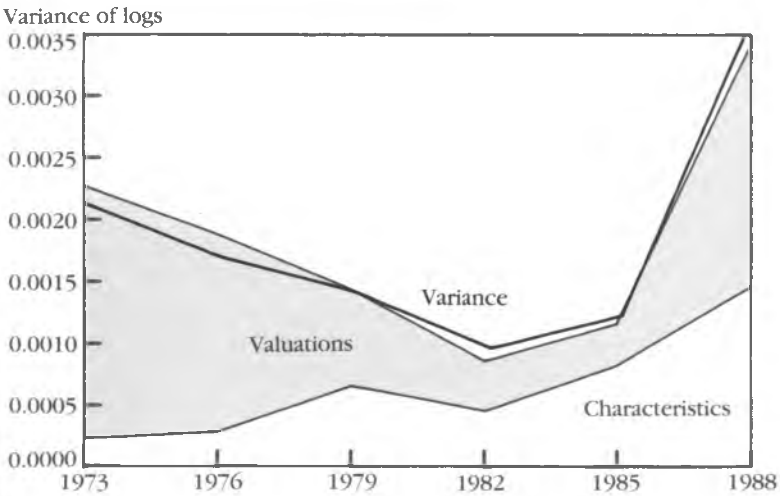
Adjusting wages by a local deflator alters these correlations. Beyond this simple statistical relationship, factors are adjusted for the degree to which higher locally adjusted wages for skills correspond to concentrations of those skills. Regional skill concentrations are fundamentally linked to the mobility decisions of workers and firms. Locally and nationally adjusted wages should result in different decompositions due to the reactions of firms and workers to wage differentials.

8 A similar decomposition is applied in Schweitzer (1993) to identify sources of earnings inequality.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

FIGURE 6

Decomposition of Nationally Adjusted Wage Variation



SOURCE: Authors' calculations.

TABLE 1

Decomposition of Nationally Adjusted Regional Wage Variation

	Total Variation	Characteristics	Valuations
		Covariance Decomposition	
1972 to 1974	0.0022	0.0002	0.0020
1975 to 1977	0.0017	0.0003	0.0016
1978 to 1980	0.0014	0.0007	0.0008
1981 to 1983	0.0009	0.0004	0.0004
1984 to 1986	0.0011	0.0008	0.0003
1987 to 1990	0.0034	0.0015	0.0020
		Variances of Factors	
1972 to 1974	0.0022	0.0006	0.0026
1975 to 1977	0.0017	0.0002	0.0016
1978 to 1980	0.0014	0.0006	0.0007
1981 to 1983	0.0009	0.0005	0.0005
1984 to 1986	0.0011	0.0008	0.0004
1987 to 1990	0.0034	0.0014	0.0020

SOURCE: Authors' calculations based on data from the U.S. Department of Labor, Bureau of Labor Statistics.

Sources of Nationally Adjusted Wage Differentials

Figure 6 addresses the question of whether the convergence/divergence pattern of regional wage differences results more from variations in labor market valuations or from variations in attribute levels (for example, the decomposition of regional wage differentials in equation [7]). The shaded area under the curve represents the portion of the variance of log wages accounted for by differences in labor market valuations. The remainder of the area under the curve is the portion of the variance explained by differences in attribute levels. In some years, the covariance decomposition terms for valuations and attributes do not add up to the total variance because of the interaction term, which is not reported. It is evident from the figure that differences in valuations follow the same U-shaped pattern as total wage variance. On the other hand, differences in workforce attributes follow a generally upward trend. This suggests that the pattern of convergence and then divergence of nationally adjusted wages results more from regional labor markets' valuing attributes differently than from an increasing dissimilarity of workers within regions. Nonetheless, regional differences associated with workforce attributes have been playing a growing role in regional wage differentials.

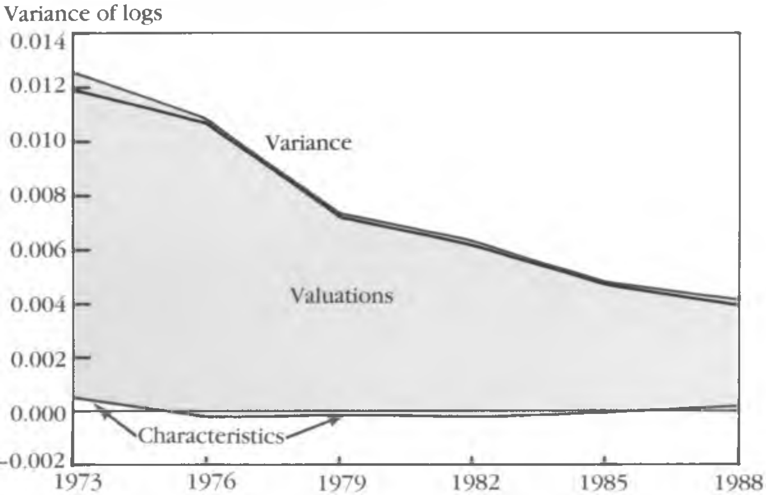
Table 1 compares the covariance decomposition results with the variances of the two significant components. In this case, the covariance between the quantity and characteristic price component is small; thus, the variances sum to approximately the total variation and are similar to the covariance decomposition terms. This confirms Eberts' (1989) results for nationally adjusted wages in a sample of the full-time metropolitan workforce.

The results in table 1 and figure 6 can be interpreted in two ways: Either incentives for firms to move toward lower-wage areas are growing, or local productive amenities are on the rise. Both conclusions hinge on our having captured the majority of worker productivity differences between regions with the worker attributes included in the wage equations. If significant productivity differences are not captured by the wage equations, and if the unobserved productivity factors have been growing nationally in value, then we could mistakenly identify productivity differentials between regions as price differences. Along these lines, Juhn, Murphy, and Pierce (1993) argue that

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

FIGURE 7

Decomposition of Locally Adjusted Wage Variation



SOURCE: Authors' calculations.

TABLE 2

Decomposition of Locally Adjusted Regional Wage Variation

	Total Variation	Characteristics	Valuations
		Covariance Decomposition	
1972 to 1974	0.0119	0.0005	0.0120
1975 to 1977	0.0107	– 0.0002	0.0110
1978 to 1980	0.0072	– 0.0001	0.0075
1981 to 1983	0.0062	– 0.0002	0.0065
1984 to 1986	0.0047	0.0000	0.0048
1987 to 1990	0.0040	0.0002	0.0039
		Variances of Factors	
1972 to 1974	0.0119	0.0006	0.0129
1975 to 1977	0.0107	0.0002	0.0116
1978 to 1980	0.0072	0.0006	0.0083
1981 to 1983	0.0062	0.0005	0.0075
1984 to 1986	0.0047	0.0008	0.0058
1987 to 1990	0.0040	0.0014	0.0052

SOURCE: Authors' calculations based on data from the U.S. Department of Labor, Bureau of Labor Statistics.

higher wage payments to unobserved skills explain the rise in total earnings inequality during the 1980s. A final caveat to our results is that the analysis does not account for fringe benefit costs. Differences in these costs between regions would of course result in a different distribution of total compensation.

Sources of Locally Adjusted Wage Differentials

While firms *might* be adjusting to these wage differentials, households *should* react to wages that reflect their cost of living. The pattern in the variance of locally adjusted wages is quite different from that of nationally adjusted wages. Instead of exhibiting a U-shaped pattern, locally adjusted wages steadily converge over the sample period (figure 7). Moreover, the dispersion of locally adjusted wages is roughly five times greater than the dispersion of nationally adjusted wages. Significantly, it is differences in labor market valuations that explain most of the total wage variance. While the dispersion in labor costs relevant to firms (nationally adjusted wages) has increased in recent years, the dispersion of regional differences in workers' returns to labor has declined.

Comparing the covariance decomposition results with simple variances indicates, in this case, that accounting for covariance between factors alters our interpretation of the components of the decline in locally adjusted wage dispersion between regions. Table 2 reveals that, unlike the nationally adjusted wage case, a significant negative covariance exists between the characteristic price component of regional wage differences and the regional distribution of attributes. This is evident both in the frequently negative quantities component and in the fact that the simple variances of the components substantially overshoot the total variances of locally adjusted wages.

Evaluating these results in terms of worker location decisions, we find that the declining differences in factor returns between regions is consistent with workers' moving to equalize labor market differences. A larger impetus for mobility is indicated by the greater wage variation between regions when cost-of-living differences are factored in. The mobility of households responding to significant, but declining, consumptive amenities in the high-price MSAs could explain this reduction in locally adjusted wage differentials between MSAs. Furthermore, the differences between locally adjusted wage

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

differentials appear to be almost purely the result of differences in valuations of labor rather than differences in labor force characteristics. These differentials could encourage significant worker movement, which could lead to rising nationally adjusted wage differences as wages are driven up in high-price areas and down in low-price areas. But it could just as well be that firms have moved to more costly areas, driving up wages, in pursuit of an amenity that has been rising in value. The unobservability of the full set of amenities, either consumptive or productive, precludes a direct test of these explanations.

V. Conclusion

The theoretically surprising fact that regional wages appeared to diverge in the 1980s does not hold up when cost-of-living differences are taken into account. Our decompositions confirm that wage differences are driven by varying returns to worker attributes rather than by regional differences in workforce characteristics. Further, the possibility is raised that workers and firms are optimizing over different value functions (nationally versus locally adjusted wages) or different local amenities. In particular, local prices, and therefore locally adjusted wages, may be more important for workers. The difference in the patterns of nationally versus locally adjusted wage differentials is consistent with a story of competing adjustments rather than of slowing adjustments.

However, other explanations are not eliminated by these results, because neither the adjustment processes nor the values of amenities have been explicitly incorporated. These shortcomings provide an obvious direction for future research. Given the limited observability of amenities, a sensible strategy would be to estimate the adjustment processes of firms and workers. This would make more explicit the link between convergence rates and differentials in the two wage series. Although our conclusion is largely descriptive, the diverse patterns in nationally versus locally adjusted wages clearly support analyzing regional wage differentials from the perspective of both employees and firms.

References

Bellante, Don. "The North–South Differential and the Migration of Heterogeneous Labor," *American Economic Review*, vol. 69 (March 1979), pp. 166–75.

Blanchard, Olivier Jean, and Lawrence F. Katz. "Regional Evolutions," *Brookings Papers on Economic Activity*, vol. 1 (1992), pp. 1–61.

Browne, Lynn E. "Shifting Regional Fortunes: The Wheel Turns," Federal Reserve Bank of Boston, *New England Economic Review*, May/June 1989, pp. 27–40.

Dickie, Mark, and Shelby Gerking. "Interregional Wage Differentials in the United States: A Survey," in Alan Schlottmann et al., eds., *Migration and Labor Market Adjustment*. Boston: Kluwer Academic Publishers, 1988.

Eberts, Randall W. "Accounting for the Recent Divergence in Regional Wage Differentials," Federal Reserve Bank of Cleveland, *Economic Review*, vol. 25, no. 3 (1989 Quarter 3), pp. 14–26.

________, and Joe A. Stone. *Wage and Employment Adjustment in Local Labor Markets*. Kalamazoo, Mich.: W.E. Upjohn Institute for Employment Research, 1992.

Farber, Stephen C., and Robert J. Newman. "Accounting for South/Non-South Real Wage Differentials and for Changes in Those Differentials over Time," *Review of Economics and Statistics*, vol. 69 (May 1987), pp. 215–23.

Gabriel, Stuart A., Janice Shack-Marquez, and William L. Wascher. "Regional Labor Markets, Cost-of-Living Differentials, and Migration," Board of Governors of the Federal Reserve System, Economic Activity Section, Working Paper No. 91, December 1988.

Gerking, Shelby, and William Weirick. "Compensating Differences and Interregional Wage Differentials," *Review of Economics and Statistics*, vol. 65 (August 1983), pp. 483–87.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

Hanoch, Giora. "An Economic Analysis of Earnings and Schooling," *Journal of Human Resources*, vol. 2 (Summer 1967), pp. 310 – 29.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy*, vol. 101, no. 3 (June 1993), pp. 410 – 42.

Mincer, Jacob. *Schooling, Experience, and Earnings*. New York: National Bureau of Economic Research, distributed by Columbia University Press, 1974.

Oaxaca, Ronald. "Male–Female Wage Differentials in Urban Labor Markets," *International Economic Review*, vol. 14 (October 1973), pp. 693 – 709.

Roback, Jennifer. "Wages, Rents, and Quality of Life," *Journal of Political Economy*, vol. 90 (December 1982), pp. 1257 – 78.

Sahling, Leonard G., and Sharon P. Smith. "Regional Wage Differentials: Has the South Risen Again?" *Review of Economics and Statistics*, vol. 65 (February 1983), pp. 131 – 35.

Schweitzer, Mark E. "Accounting for Earnings Inequality in a Diverse Workforce," Federal Reserve Bank of Cleveland, Working Paper 9314, December 1993.

Shryock, Henry S., and Jacob S. Siegel. *The Methods and Materials of Demography*, vol. 1. Washington, D.C.: U.S. Department of Commerce, Bureau of the Census, 1971.

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

Economic Review

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

Second Quarter Working Papers

Current *Working Papers* of the Cleveland Federal Reserve Bank are listed in each quarterly issue of the *Economic Review.* Copies of specific papers may be requested by completing and mailing the attached form below.

Single copies of individual papers will be sent free of charge to those who request them. A mailing list service for personal subscribers, however, is not available.

Institutional subscribers, such as libraries and other organizations, will be placed on a mailing list upon request and will automatically receive *Working Papers* as they are published.

■ **9405**
The Federal Reserve Board before Marriner Eccles
by Walker F. Todd

■ **9406**
Optimal Fiscal Policy when Public Capital Is Productive: A Business Cycle Perspective
by Kevin J. Lansing

■ **9407**
Anticipating Bailouts: The Incentive–Conflict Model and the Collapse of the Ohio Deposit Guarantee Fund
by Ramon P. DeGennaro and James B. Thomson

Please complete and detach the form below and mail to:
Research Department
Federal Reserve Bank of Cleveland
P.O. Box 6387
Cleveland, Ohio 44101

Check item(s) requested

Please send the following Working Paper(s):

☐ **9405** ☐ **9406** ☐ **9407**

Send to:
Please print

Name

Address

City State Zip

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis

Economic Commentary

Enterprise Liability: A Prescription for Health Care Reform?
by Charles T. Carlstrom
July 1, 1993

The Evolving Loan Sales Market
by Joseph G. Haubrich and James B. Thomson
July 15, 1993

Free Markets and Price Stability: Opportunities for Mexico
by Jerry L. Jordan
August 1, 1993

Assessing Real Interest Rates
by John B. Carlson
August 15, 1993

Airline Deregulation: Is It Time to Finish the Job?
by Paul W. Bauer and Ian Gale
September 1, 1993

The Decline in U.S. Saving Rates: A Cause for Concern?
by Jagadeesh Gokhale
September 15, 1993

Credibility Begins with a Clear Commitment to Price Stability
by Jerry L. Jordan
October 1, 1993

The Budget Reconciliation Act of 1993: A Summary Report
by David Altig and Jagadeesh Gokhale
October 15, 1993

Making the SAIF Safe for Taxpayers
by William P. Osterberg and James B. Thomson
November 1, 1993

Community Lending and Economic Development
by Jerry L. Jordan
November 15, 1993

Replacing Reserve Requirements
by E.J. Stevens
December 1, 1993

Long-Term Health Care: Is Social Insurance Desirable?
by Jagadeesh Gokhale and Lydia K. Leovic
December 15, 1993

Monetary Policy and Inflation: 1993 in Perspective
by Gregory A. Bauer and John B. Carlson
January 1, 1994

Report of the Fourth District Economists' Roundtable
by Michael F. Bryan and John B. Martin
January 15, 1994

Are Service-Sector Jobs Inferior?
by Max Dupuy and Mark E. Schweitzer
February 1, 1994

The National Depositor Preference Law
by James B. Thomson
February 15, 1994

Issues in CRA Reform
by Mark S. Sniderman
March 1, 1994

Back to the Future: A View of Prospective Deficits through the Prism of the Past
by David Altig and Jagadeesh Gokhale
March 15, 1994

Central Bank Independence
by Owen F. Humpage
April 1, 1994

Health Care Reform from a Generational Perspective
by David Altig and Jagadeesh Gokhale
April 15, 1994

Lessons from the Collapse of Three State-Chartered Private Deposit Insurance Funds
by Walker F. Todd
May 1, 1994

Assessing Progress toward Price Stability: Looking Forward and Looking Backward
by John B. Carlson
May 15, 1994

The Government's Role in the Health Care Industry: Past, Present, and Future
by Charles T. Carlstrom
June 1, 1994

Are the Japanese to Blame for Our Trade Deficit?
by Owen F. Humpage
June 15, 1994

A Beginner's Guide to the U.S. Payments System
by Paul W. Bauer
July 1, 1994

Digitized for FRASER
http://fraser.stlouisfed.org/
Federal Reserve Bank of St. Louis