Federal Reserve Bank
of Chicago

Second Quarter 2001

Economic perspectives

Economic perspectives

Economic Perspectives is published by the Research Department of the Federal Reserve Bank of Chicago. The views expressed are the authors' and do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System.

Single-copy subscriptions are available free of charge. Please send requests for single- and multiple-copy subscriptions, back issues, and address changes to the Public Information Center, Federal Reserve Bank of Chicago, P.O. Box 834, Chicago, Illinois 60690-0834, telephone 312-322-5111 or fax 312-322-5515.

Economic Perspectives and other Bank publications are available on the World Wide Web at http://www.chicagofed.org.

chicagofed.org

ISSN 0164-0682

Contents

Second Quarter 2001, Volume XXV, Issue 2

Are the large central cities of the Midwest reviving?

William A. Testa

Introduction and summary

Most central cities of the large metropolitan areas of the Midwest showed signs of improvement during the 1990s compared with the previous two decades, according to such broad measures as population, employment, unemployment, and income. If such gains can be sustained, it will be welcome news for households residing in central cities who experienced erosion of their income and tax base during the second half of the 20th century. Such gains might also provide important evidence of the results of the recent policies of big city mayors, who have been very active in both improving the quality of urban life—through transportation, crime, and school reform initiatives—and engaging in economic development initiatives—such as work force training and rebuilding city infrastructure. In this article, I analyze broad measures of 11 central city economies since 1970 to assess whether there has been any underlying structural improvement in big city performance beyond the effects of the general U.S. and regional economic expansion.

I relate each city's performance to that of its surrounding suburban areas. In this way, I can control for many factors that may be peculiar to a given metropolitan area—such as a change in the performance of an area's key industry and overall economy or its location on a particular interstate highway. Within this framework, I ask whether the city's share of metropolitan population and employment is growing over time, or at least whether its loss of share is abating, and whether other performance measures such as household income and unemployment rate are improving in the city relative to its suburbs. Such a standard for improvement may be stringent. Most of the 11 large central cities of the Midwest have fixed boundaries; they are unable to annex land to accommodate population growth in the metropolitan area, while the surrounding suburban areas are able to do so.

I find that, on average, the population of the 11 cities almost stabilized in the 1990s, a marked improvement compared with the 1970s. In part, however, it appears that central city population recovery largely reflects buoyant regionwide recovery rather than structural change; central cities continue to lose share of population to their suburbs. However, my analysis of total permits to construct residential housing units indicates that, although cities continued to lose ground to their suburbs in the 1990s relative to the 1980s, single-family construction showed the opposite trend, perhaps reflecting the much-touted recovery of central cities as a livable place for families. So too, the Midwest's economic recovery of the 1990s has lifted labor force participation and income in both city and suburb. Furthermore, tightening labor markets in the 1990s clearly narrowed the gap between suburban unemployment rates and those of the city, although the low ratios of household income in cities versus their suburbs have not improved.

It appears that city residents continue to look to the periphery of metropolitan areas to earn their income. At least through 1997, job sites continued to decentralize from the center of the metropolitan area. Overall, I conclude that, although there are several individual instances to the contrary, central cities in the Midwest continued to struggle to keep pace with their suburbs in the 1990s in terms of job growth and economic development. Nonetheless, there are some positive indications for the future, and it is quite evident that the large central cities of the Midwest have

William A. Testa is vice president and director of regional programs in the Economic Research Department of the Federal Reserve Bank of Chicago. *The author thanks Margrethe Krontoft and Loula Merkel for research assistance.*

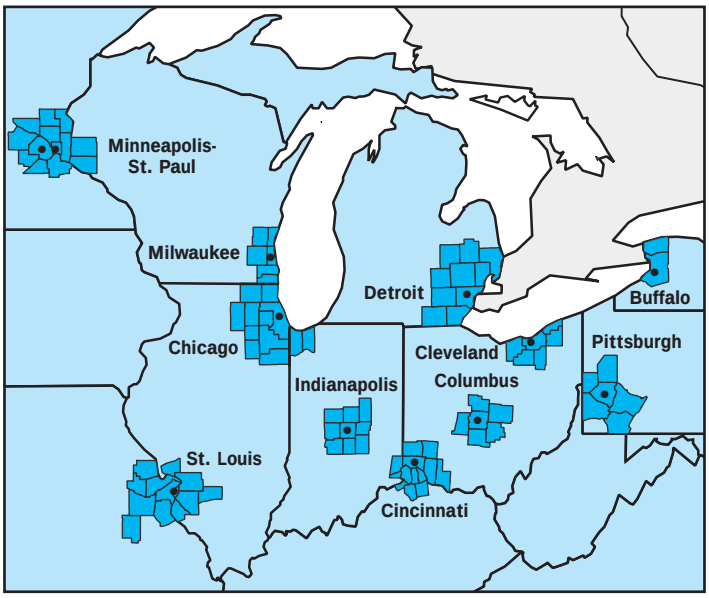
shared in the bounty of the general economic recovery.

Are cities gaining population and housing?

In the U.S. and in most developed countries, households exercise choice in where to locate their residences. Accordingly, population growth is a frequently examined indicator of the health and attractiveness of a locale. In the 11 metropolitan areas chosen for this article, the central cities continue to comprise a major, though declining, share of the populations of their respective metropolitan areas (see figure 1 and table 1). According to recent data from the Bureau of the Census, these cities comprised 28 percent of their metropolitan statistical area (MSA) population in 2000. Combined, the cities represented a 50 percent share of the population of the metropolitan area at mid-century and a 55 percent share in 1900 (table 1).[1]

FIGURE 1

Major midwestern cities and their metropolitan areas



Source: U.S. Department of Commerce, Bureau of the Census.

How did these cities fare during the 1990s in comparison to the 1980s? Looking first at population growth in central cities, we see that six cities experienced an improvement in their average annual growth rate of population in the 1990s—Chicago, Cleveland, Detroit, Indianapolis, Minneapolis-St. Paul, and Pittsburgh (table 2). Of these, only Chicago, Indianapolis, and Minneapolis-St. Paul actually grew; the population of Cleveland, Detroit, and Pittsburgh declined more slowly than in the previous decade. The population changes in these six cities combined were sufficient to offset the deterioration in the other five central cities, so that the average growth of the total city population registered an improvement from the 1980s to the 1990s, wherein the annual growth rate climbed from –1.3 percent per year to stable population on average. An unweighted average, whereby each city is given equal weight, shows that average annual population growth improved slightly from a loss of .5 percent per year over 1980–90 in comparison to a loss of .2 percent per year over the 1990–2000 period.

In comparing the 1980s to the 1990s, the improvements are more widespread. All 11 central cities experienced improvements in population change. This is not too surprising since overall population growth of the metropolitan areas that overlie the central cities accelerated in the 1990s, supported by the economic turnaround in the Midwest. Migration out of the Midwest has slowed to a trickle in recent years, and population growth in the 11 sample metropolitan areas accelerated from .2 percent per year in the 1980s to .7 percent per year in the 1990s. But was there a shift in residential preferences between city

TABLE 1

Eleven Great Lakes cities

	2000 population		City as share of MSA		
	City	MSA	1900	1950	2000
	(thousands)		*(percent)*		
Buffalo	293	1,170	69.3	53.3	25.0
Chicago	2,896	8,273	74.9	62.4	35.0
Cincinnati	331	1,646	44.0	41.3	20.1
Cleveland	478	2,251	54.2	41.0	21.3
Columbus	711	1,540	39.3	53.1	46.2
Detroit	951	4,442	40.8	49.8	21.4
Indianapolis	782	1,607	39.2	51.4	48.6
Milwaukee	597	1,501	63.3	56.7	39.8
Minneapolis-St. Paul	670	2,969	60.5	64.3	22.6
Pittsburgh	335	2,359	25.7	27.1	14.2
St. Louis	348	2,604	60.7	46.6	13.4
All 11 cities	**8,392**	**30,361**	**54.7**	**50.4**	**27.6**

Notes: MSA is metropolitan statistical area. MSA reflects 1998 definition for all years.
Source: U.S. Department of Commerce, Bureau of the Census, various years.

TABLE 2

Average annual change in population and share of MSA

	Population			Share of MSA		
	1970–80	1980–90	1990–2000	1970–80	1980–90	1990–2000
	(percent)			*(percentage points)*		
Buffalo	–2.3	–0.8	–1.1	–1.6	–0.4	–0.9
Chicago	–1.1	–0.7	0.4	–1.2	–0.9	–0.7
Cincinnati	–1.5	–0.6	–0.9	–1.8	–1.0	–1.6
Cleveland	–2.4	–1.2	–0.5	–1.9	–1.0	–0.7
Columbus	0.5	1.2	1.2	–0.3	0.1	–0.2
Detroit	–2.0	–1.5	–0.7	–2.0	–1.3	–1.2
Indianapolis	–0.6	0.5	0.7	–1.0	–0.1	–0.8
Milwaukee	–1.1	–0.1	–0.5	–1.1	–0.4	–0.9
Minneapolis-St. Paul	–1.4	0.0	0.5	–2.1	–1.3	–1.5
Pittsburgh	–1.8	–1.3	–1.0	–1.5	–0.6	–0.8
St. Louis	–2.7	–0.6	–1.2	–2.6	–1.6	–1.6
Weighted avg.	–1.4	–1.3	0.0	–1.4	–0.8	–0.7
Unweighted avg.	–1.5	–0.5	–0.2	–1.5	–0.8	–1.0
11 MSAs	0.0	0.2	0.7	n.a.	n.a.	n.a.
U.S.	1.4	1.2	1.0	n.a.	n.a.	n.a.

Notes: n.a. indicates not applicable. MSA is metropolitan statistical area.
Source: U.S. Department of Commerce, Bureau of the Census, various years.

and suburbs in the 1990s? Here again we see that most central cities are indeed moving in a positive direction in comparison to the 1970s (table 2). All appear to be either experiencing a deceleration in loss of share or an acceleration in gains. However, in the aggregate a modest deterioration occurred from the 1980s to 1990s as measured by the unweighted average. Buffalo, Cincinnati, Columbus, Indianapolis, Milwaukee, Minneapolis-St. Paul, and Pittsburgh saw an increased rate in the erosion of population share to their suburbs.

In assessing the importance of these population losses in central cities, it is important to note that the municipal boundaries of the cities have remained essentially fixed while those of their metropolitan areas have expanded to accommodate growth in households and rising demand for housing and land. The rising demand for space means, for example, that there will be a growing share of population in those parts of the metropolitan area where land area can expand. In point of fact, the boundaries of large midwestern cities have not grown much. Notable exceptions to stagnant city boundaries are Columbus, Ohio, which has used its strategic assets of water and sewerage treatment capacity to induce annexation of neighboring development; Indianapolis, which became roughly coincident with its surrounding county government all in one fell swoop in the 1970s; and Milwaukee, which undertook an aggressive, but short-lived, annexation policy during the 1950s (table 3). The remaining eight cities taken together expanded their land area by only 3.7 percent from 1950 to 1990.

The overall population densities of metropolitan areas have been falling steeply since the early decades of the twentieth century, thereby spreading out existing population. Households tend to live today in a fashion that consumes more housing—both land and structure—than earlier in the century. Accordingly, even had no further population increase taken place in metropolitan areas, households would have jumped the fixed city boundary in achieving lower densities of living (and working), thereby reducing population of central cities. The trend toward declining densities in central cities can be seen between 1920 and 1990 (table 4). For all 11 cities taken together, and not adjusting for changing city boundaries and land area, average density declined by almost one-half over the period. Even if we exclude Indianapolis, Columbus, and Milwaukee—whose boundaries were highly expansionary—average city density declined by approximately one-half over this period. The second two columns of table 4 measure the rate at which population density in the entire metropolitan area falls for every mile of distance from the center of the city.

TABLE 3

Land area (square miles) and density (population per square mile)

	1910	1920	1930	1940	1950	1960	1970	1980	1990
Buffalo									
Land area	38.7	38.9	38.9	39.4	39.4	39.4	41.3	41.8	40.6
Density	10,949	13,028	14,732	14,617	14,724	13,522	11,205	8,561	8,082
Chicago									
Land area	185.1	192.8	201.9	206.7	207.5	224.2	222.6	228.1	227.2
Density	11,806	14,013	16,723	16,434	17,450	15,836	15,126	13,174	12,252
Cincinnati									
Land area	49.8	71.1	71.4	72.4	75.1	77.3	78.1	78.1	77.2
Density	7,301	5,643	6,319	6,293	6,711	6,501	5,794	4,935	4,716
Cleveland									
Land area	45.6	56.4	70.8	73.1	75.0	81.2	75.9	79.0	77.0
Density	12,295	14,128	12,718	12,016	12,197	10,789	9,893	7,264	6,566
Columbus									
Land area	20.3	22.6	38.5	39.0	39.4	89.0	134.6	180.9	190.9
Density	8,941	10,488	7,547	7,848	9,541	5,296	4,009	3,123	3,315
Detroit									
Land area	40.8	77.9	137.9	137.9	139.6	139.6	138.0	135.6	138.7
Density	11,416	12,748	11,375	11,773	13,249	11,964	10,953	8,874	7,411
Indianapolis									
Land area	33.0	43.6	54.2	53.6	55.2	71.2	379.4	352.0	361.7
Density	7,080	7,206	6,719	7,220	7,739	6,689	1,963	1,991	2,022
Milwaukee									
Land area	22.8	25.3	41.1	43.4	50.0	91.1	95.0	95.8	96.1
Density	16,397	18,069	14,069	13,536	12,748	8,137	7,548	6,641	6,536
Minneapolis-St. Paul									
Land area	102.3	101.9	107.6	106.0	106.0	108.7	107.3	107.5	107.7
Density	5,045	6,038	6,840	7,359	7,859	7,326	6,937	5,964	5,948
Pittsburgh									
Land area	41.4	39.9	51.3	52.1	54.2	54.1	55.2	55.4	55.6
Density	12,896	14,745	13,057	12,892	12,487	11,171	9,422	7,652	6,653
St. Louis									
Land area	61.4	61.0	61.0	61.0	61.0	61.0	61.2	61.4	61.9
Density	11,189	12,670	13,475	13,378	14,046	12,296	10,167	7,379	6,408
All 11 cities									
Land area	641.2	731.4	874.6	884.6	902.4	1,036.8	1,388.6	1,415.6	1,434.6
Density	10,176	11,464	11,812	11,845	12,496	10,582	7,513	6,319	5,862

Source: U.S. Department of Commerce, Bureau of the Census, various years.

From this we see that population densities have been declining both within and outside of central cities. What are the underlying reasons for these falling densities?

Changing technologies and standards of living are generally thought to have given rise to decisions of city residents to decentralize. Significant technological forces spurring lower-density living and working are described as pervasive by urban analysts and are reflected in the trend of suburbanization around the world.[2] Rising household incomes pushed families to desire more housing and land, trading off longer working commutes to the central city for more (and distant) housing where land was cheaper. Falling automobile prices and better highways in the early twentieth century lent a further impetus to suburban living. Meanwhile, on the production and employment side, there was also strong impetus to decentralization. Highways freed factories from their ties to water ports, railroads, and rail spurs. Intermediate goods could be shipped in from afar on trucks, and final goods sent out the same way. So too, workers at inner-city factories increasingly gave way to machinery, and those few workers no longer needed to walk or take a streetcar to the factory site. With assembly-line production assisted by electric tools and conveyor belts, multistory factories converted or moved to sprawling land intensive one-story buildings. And why not build those low-slung modern factories where land was

TABLE 4

Population density

City	Density (population per square mile)		Percent falloff in density per mile from city center	
	1920	1990	1920	1990
Buffalo	13,028	8,082	0.15	0.13
Chicago	14,013	12,252	0.15	0.09
Cincinnati	5,643	4,716	0.23	0.13
Cleveland	14,128	6,566	0.22	0.11
Columbus	10,488	3,315	0.22	0.12
Detroit	12,748	7,411	0.19	0.11
Indianapolis	7,206	2,022	0.24	0.07
Milwaukee	18,069	6,536	0.31	0.16
Minneapolis-St. Paul	6,038	5,948	0.18	0.11
Pittsburgh	14,745	6,653	0.17	0.12
St. Louis	12,670	6,408	0.22	0.11
All 11 cities	**11,707**	**6,355**	**0.21**	**0.11**
Standard deviation			0.05	0.02

Source: Author's calculations based on decennial census data.

cheaper and transportation/warehousing more commodious, that is, far distant from the city center. In the latter part of the twentieth century, job location followed population in suburbanizing, so much so that metropolitan areas can often be characterized as containing several large employment centers dispersed throughout the metropolitan area.[3]

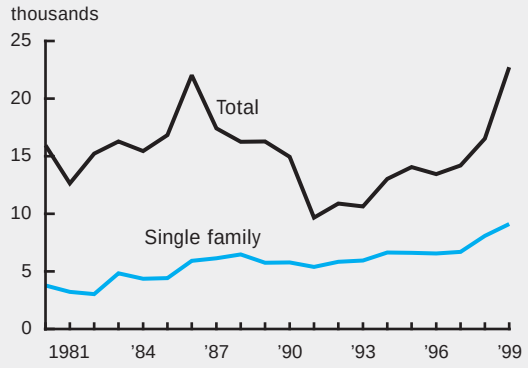
This portrayal implies that midwestern cities may now be in the process of lowering or equalizing their densities to match their surrounding suburbs. Adjustment to lower densities cannot take place instantaneously. Both residential and nonresidential capital in the form of housing, commercial buildings, and public infrastructure are far from perfectly malleable.[4] Even as demand favors less dense residential and commercial space, rents will tend to fall below the costs of new construction, thereby forestalling de-concentration pending the depreciation of the stock of existing buildings. Thus, some observers propose that city decline is partly a transitory and delayed adjustment of density to new technology, which further implies that the cities' population decline will bottom out at some point when an equilibrium density is achieved. The fact that the technologies of overland transportation and industrial production are no longer making those significant technological leaps that have lowered preferred density gives rise to some optimism that city population decline may soon bottom out to an equilibrium state of land use density with the surrounding metropolitan area.

On the other hand, some observers suggest that tastes may change back toward a preference for residential living in a more compact form. One school of thought called "new urbanism" is now promoting higher density residential lifestyles within walking distance to shopping, entertainment, and public transportation. In fact, observers have reported on the pickup in the pace of residential building in some central cities in the late 1990s. This phenomenon has been attributed to a revived interest in city living by both young and old, but mostly childless, households. An expected demographic movement toward larger numbers of childless households as baby-boomers pass their child-rearing years may presage a continued revival of interest in city living. Meanwhile, in attempting to retain and attract families, central cities such as Milwaukee, Cleveland, Detroit, and Chicago have launched ambitious and innovative initiatives to improve their public school systems.

As to hard evidence of growth in housing activity, municipal governments typically report permits that are filed in advance of construction (and conversion) of new housing units. An unknown portion of these permits are not acted on, and there is no timely data source available on abandonment or tear-downs with which to assess changes to the overall net stock of housing. Nonetheless, these data do indicate the expected and planned level of new residential construction activity. Figure 2 shows the pace of building permits of residential units back to 1980, and there is clearly steady growth in the 1990s, with a marked acceleration in the past two to three years. Single-family home building growth is especially steady in its upward climb, with both total (and multi-unit) housing being much more volatile. However, in the context of business cycle movements, the recent rise in building is somewhat less impressive; most midwestern cities are only now reaching the levels of residential building activity that were previously attained in the mid to late 1980s. For all 11 cities combined, the number of residential permits issues for the last five years of the 1990s reached only 90.6 percent of the levels for the late 1980s (table 5, column 2). However, the data are more sanguine for single-family housing permits. In the last five years of the 1990s, single-family permits were taken out at a much more rapid clip in central cities compared with the last five years of the 1980s (table 5, column 5). In fact, the improvement in the pace of

FIGURE 2

Authorized residential units
(11 midwestern cities)



Sources: U.S. Department of Commerce, Bureau of the Census, and U.S. Department of Labor, Bureau of Labor Statistics, various years, *Current Population Survey.*

permits for single-family housing in cities even compares favorably with the suburban areas of MSAs.

Are city residents doing better?

We have seen that population and housing growth, or a slowing in the pace of decline, may be a sign of city revival as households increasingly come to view the city favorably and choose to live there. However, because technologies have universally changed living and working for the better, those who choose to remain in the city may also be better off. Apart from geographic growth measures, then, what are the more direct measures of the well-being of city residents that we can compare with suburban counterparts? Both average household income and the unemployment rate are powerful and widely accepted measures of well-being. Household income estimates for cities and their surrounding metropolitan areas can be constructed from sample data collected annually by the Bureau of the Census and the Bureau of Labor Statistics in their *Current Population Survey*. A second measure reflects the degree to which city residents have access to opportunities to participate in the work force. Local unemployment rates are constructed through sampling of the members of working age households by the Bureau of Labor Statistics in cooperation with state employment agencies.

These indicators show absolute improvements for city residents in the 1990s (figure 3). Unemployment rates averaged over the central cities peaked at over 15 percent in the early 1980s, and have since declined to a recent level of approximately 6 percent for workers aged 16 years and older. Similarly for real average household income (deflated by the Consumer Price Index calculated for all urban areas), the

TABLE 5

Residential permits (ratios x 100)

	Total residential units			Single family units		
	1990–94 / 1980–84	1995–99 / 1985–89	1990s / 1980s	1990–94 / 1980–84	1995–99 / 1985–89	1990s / 1980s
Buffalo	239.4	202.1	216.2	234.0	76.8	118.3
Chicago	70.0	129.2	98.4	196.9	167.3	178.7
Cincinnati	101.6	83.7	94.9	530.5	133.2	244.9
Cleveland	90.5	143.7	120.3	738.6	803.3	785.4
Columbus	108.7	77.1	88.5	133.2	95.0	110.3
Detroit	47.3	96.0	62.1	51.7	479.8	125.3
Indianapolis	93.4	86.4	89.2	200.1	140.8	161.4
Milwaukee	59.8	58.6	59.2	41.9	75.1	52.1
Minneapolis-St. Paul	24.1	94.9	45.2	51.2	227.4	102.6
Pittsburgh	30.4	76.7	49.6	59.4	122.5	86.9
St. Louis	12.6	71.6	35.3	143.4	129.7	132.9
All 11 cities	**78.3**	**90.6**	**85.0**	**154.0**	**129.3**	**139.2**
U.S.	**88.6**	**95.4**	**92.3**	**123.8**	**111.6**	**116.8**

Note: Ratios of earlier versus later five-year period or decade.
Source: U.S. Department of Commerce, Bureau of the Census, various years.

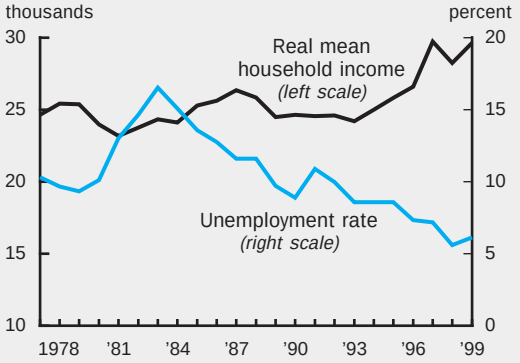
trend was for sideways movement from the late 1970s up until the early 1990s, from which point the current expansion has lifted mean incomes by 15 percent to 20 percent. There is no question, then, that the 1990s have on average lifted the fortunes of city residents.

How have city residents fared versus suburban residents? Average household incomes in comparison to suburban counterparts have not changed appreciably from the 1980s (table 6). Again, we can look at these over comparable periods of the 1980s and 1990s. Interestingly, it appears that city incomes are somewhat countercyclical—really less procyclical—compared with the suburbs; the income ratio of city to suburb tends to climb during contractions and fall during expansions (figure 4). Perhaps one explanation is that a greater proportion of city residents depend on fixed income streams from pensions and government income support programs than their suburban counterparts. Such income streams are less likely to evaporate during a downturn. In any event, the relative income of city residents versus suburbs has not improved from the latter 1980s, which was a similar business cycle period to the latter 1990s.[5] More formal trend analysis (not reported) using ordinary least squares (OLS) multiple regression does not suggest that the procyclicality of the suburb to city ratio is statistically significant. Moreover, a binary variable for 1990–95 and one for 1996–99 suggest that the suburb to city ratio of mean household income widened during the booming 1990s. Real household income has risen in both city and suburb alike, but more so for suburban households.

What do unemployment rates say about the economic well being of city residents? Currently, there is little doubt that the Midwest's tight labor markets

FIGURE 3

Unemployment rates and real income
(11 midwestern cities)



Sources: U.S. Department of Commerce, Bureau of the Census, and U.S. Department of Labor, Bureau of Labor Statistics, various years, *Current Population Survey.*

are lifting the employment rates of city populations. Though these are an imperfect measure of employment participation, unemployment rates in both city and suburb alike are the lowest seen in 30 years. To assess whether cities are coming back within the context of their surrounding regions, I focus on explaining the difference between each city's unemployment rate minus the adjacent suburban area's unemployment rate (in March of each year) for adults aged 16 years and over. Over a combined sample of each of the years from 1977 to 1999, I regressed this unemployment rate gap against each city's overarching MSA unemployment rate (see box 1). This MSA unemployment rate—an independent variable in the regression—accounts

TABLE 6

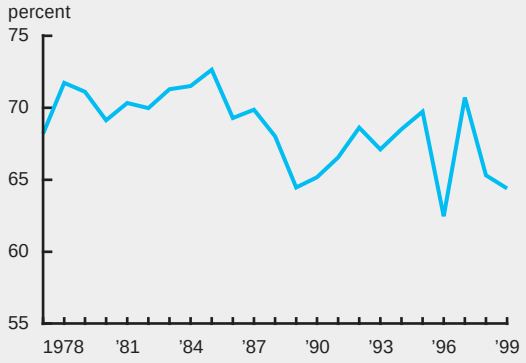
Average city to suburb ratios of mean income

	1960	1970	1980–85	1986–89	1990–95	1996–99
Buffalo	0.81	0.69	0.61	0.63	0.60	0.67
Chicago	0.80	0.71	0.67	0.66	0.65	0.63
Cincinnati	0.83	0.76	0.70	0.74	0.84	1.09
Cleveland	0.74	0.64	0.72	0.56	0.51	0.50
Columbus	0.77	0.78	0.75	0.74	0.75	0.70
Detroit	0.81	0.69	0.62	0.58	0.55	0.50
Indianapolis	0.79	1.01	1.00	0.88	0.81	0.63
Milwaukee	0.86	0.71	0.64	0.69	0.67	0.61
Minneapolis-St. Paul	0.87	0.70	0.73	0.66	0.78	0.69
Pittsburgh	0.88	0.82	0.82	0.81	0.79	0.92
St. Louis	0.76	0.67	0.58	0.57	0.50	0.64
All 11 cities	**n.a.**	**0.74**	**0.71**	**0.68**	**0.67**	**0.66**

Note: n.a. indicates not available. 1960 data represent median family income for central cities and urban fringes of urbanized areas.
Sources: 1960 and 1970 data are from the decennial census. All other data are from the March CPS.

FIGURE 4

City/suburbs mean income
(11 midwestern cities)



Sources: U.S. Department of Commerce, Bureau of the Census and U.S. Department of Labor, Bureau of Labor Statistics, various years, *Current Population Survey*.

for the specific point of the business cycle for each particular metropolitan area, as well as accounting for the overall MSA-specific labor market condition. As an estimation strategy, I include so-called fixed effects—that is, a binary or "shift" variable for each metropolitan area—in the regression equation to account for differences in each individual region's industry and work force composition.

In reviewing the regression results, I find clear evidence that unemployment rates in the city gained on the suburbs during the very strong labor markets of the 1990s (table 7). The estimated effects of the shift variables for the 1990s and for the 1995–99 period indicate that the gap has narrowed in unemployment rate between suburb and city. Lower metropolitan unemployment rates during the 1990s have tended to dampen city unemployment rates even more. In the event that the current tight labor markets persist, as the ongoing trend toward slower growth of the U.S. work force suggests, the city's working age residents may continue to enjoy abundant employment opportunities.

Are cities a better workplace?

The location of employment is an important indicator of a city's economic base. For one reason, such employment usually reflects the richness of the taxable base from which municipal and school district governments can raise revenues to provide services to city residents. Secondly, such jobs importantly reflect employment opportunities to residents that are accessible and proximate—jobs from which city households can generate their own wealth and income. How, then, are the large midwestern cities faring as sites for employment, especially in relation to their suburbs?

Jobs have been suburbanizing at a phenomenal pace in recent decades, so much so that the "reverse commute" from city to suburbs now rivals that of suburb to city. As of the 1960 *Census of Population*, the net flow of workers to central city job sites (on a population-adjusted basis) clearly favored the city; 36.6 percent of employed suburban residents worked in the 11 major central cities, while only 9.4 percent of city residents worked in their suburbs. This has changed dramatically. By the 1990 census, 26.2 percent of city residents commuted outward, while 28.4 percent of suburbanites headed for city job sites.[6]

Data covering jobs located in central cities is sparser than that for population, income, and employment. Indeed, the decennial census provides our only intermittent glimpse of the evolution of jobs in central cities. On a timely and consistent historical basis,

BOX 1

Analyzing MSA growth trends by analyzing employment rates

To formally test for a changing trend in the unemployment rate of central cities versus their own suburbs, I use an ordinary least squares regression equation, with the difference in city minus suburban unemployment rate as the dependent variable to be explained. I use annual observations for each of the 11 cities for each year from 1977 to 1999 as the dependent variable. The regression equation becomes

$$URDIF_{it} = \beta_i P_i + \beta_2 UR_{it} + \beta_3 Y_t + \varepsilon_t,$$

where $URDIF_{it}$ represents the difference of the city's unemployment rate in metropolitan area i from the suburban area's unemployment rate in the same region at time t. Coefficients β_i ($i = 1, 2, 3, \ldots 11$) are estimated for each metropolitan region i observed as P_i. Since these observations are loaded as zero or one (indicating place), the coefficients β_i act as shifters to pick up region-specific differences in suburban minus city labor markets. The effect on *URDIF* of each metropolitan area's overall labor market condition is estimated by the coefficient, β_2, acting through UR_{it}, the overall metropolitan area unemployment rate, which is observed to vary across time t and place i. The coefficient β_3 is the estimated effect of the particular year acting on the observations Y_t observed as period 1990–99 or 1995–99, respectively. Since these observations as loaded as zero or one (for the specified period), the coefficient reflects another shifter, testing whether *URDIF* has shifted during these periods relative to previous years 1977–89.

TABLE 7

Effect of place and time on city versus suburban unemployment

Independent variable	(Dependent variable: $(UR_{city} - UR_{subs})$, 1977–99)					
Buffalo	2.54 (2.1)*	1.89 (1.6)	5.52 (7.5)*	5.12 (7.0)*	2.54 (2.0)*	1.86 (1.6)
Chicago	3.89 (3.6)*	3.29 (3.1)*	6.41 (8.7)*	6.02 (8.2)*	3.88 (3.4)*	3.26 (3.1)*
Cincinnati	2.33 (2.1)*	1.73 (1.6)	4.89 (6.6)*	4.50 (6.1)*	2.31 (2.0)*	1.66 (1.5)
Cleveland	6.27 (6.0)*	5.70 (5.6)*	8.62 (11.6)*	8.23 (11.2)*	6.26 (5.6)*	5.65 (5.5)*
Columbus	2.19 (2.2)*	1.67 (1.8)*	4.22 (5.7)*	3.83 (5.2)*	2.18 (2.1)*	1.63 (1.7)*
Detroit	8.39 (6.9)*	7.73 (6.6)*	11.45 (15.5)*	11.05 (15.0)*	8.39 (6.6)*	7.69 (6.4)*
Indianapolis	0.94 (0.9)	0.39 (0.4)	3.15 (4.1)*	2.74 (3.6)*	0.94 (0.9)	0.35 (0.4)
Milwaukee	3.40 (3.4)*	2.85 (2.9)*	5.58 (7.5)*	5.19 (7.0)*	3.39 (3.1)*	2.82 (2.8)*
Minneapolis-St. Paul	1.03 (1.1)	0.54 (0.6)	2.82 (3.8)*	2.42 (3.3)*	1.02 (1.0)	0.50 (0.5)
Pittsburgh	1.42 (1.3)	0.81 (0.7)	4.09 (5.5)*	3.70 (5.0)*	1.41 (1.2)	0.78 (0.7)
St. Louis	4.08 (3.9)*	3.52 (3.5)*	6.38 (8.6)*	5.99 (8.1)*	4.07 (3.8)*	3.49 (3.5)*
Unemployment rate in metro area	0.29 (3.1)*	0.33 (3.6)*	—	—	0.29 (2.8)*	0.34 (3.5)*
Year shifter						
1990–99	–1.35(–2.6)*	—	–2.22 (–5.0)*	—	–1.32 (–1.0)	—
1995–99	—	–1.15 (–1.9)*	—	–2.26 (–4.3)*	—	–0.82(–0.5)
Interaction of place and time						
1990–99	—	—	—	—	–0.01 (0)	—
1995–99	—	—	—	—	—	–0.07(–0.2)
R^2	0.73	0.73	0.72	0.72	0.73	0.73
Durbin–Watson	1.91	1.92	1.85	1.85	1.91	1.92

* Denotes significance at 90 percent level.
Notes: T-stats in parentheses; data not available for 1994 (all cities) and Indianapolis for 1989.
Source: U.S. Department of Commerce, Bureau of the Census, various years, CPS supplement, March.

there has been no data series collected to reflect city boundaries. For this reason, it is difficult to measure the decentralization of job sites into the 1990s and to analyze it in the context of previous decades. As a substitute, I use the comprehensive annual estimates of employment by location at the *county* level of geography from the Bureau of Economic Analysis, which are reported back to 1969.[7] I can use these data to compare central county data trends with those of surrounding suburbs to assess the progress of central areas as job sites in the 1990s. To corroborate my findings, I piece together job data covering many (but not all) individual industries in the city versus the suburbs, as reported in various census reports of industry sectors from the U.S. Census Bureau. These, admittedly incomplete, data tend to corroborate the assertion that, while conditions have definitely improved, there is little in the way of structural or comparative improvement of cities in relation to suburban growth.

Beginning with the county data, the pattern that emerges is much like that of population trends. As shown in table 8, the average annual employment growth rate in central counties improved modestly from .7 percent per year from the 1969–79 period to .9 percent during the 1979–89 period. Perhaps that improvement is not too surprising given the propensity for there to be a mutual attraction between job location and residential location. However, job growth showed no improvement from the decade of the 1980s to the decade of the 1990s (up through 1998). Taken together, employment growth remained constant at .9 percent per year; taken as a group with each observation given equal weight, growth deteriorated from 1.2 percent per year to 1.0 annual growth in the 1990s.

Has there been any underlying structural improvement in the trends for central counties? When I compare the performance of central counties to their surrounding counties, I find that little if any overall improvement has taken place. The 1980s display an easing of the rate of loss in comparison to the 1970s (table 8, columns 4, 5, and 6). Yet, on average, the 1990s appear to have experienced acceleration in share loss from the 1980s, and in fact to have performed no better and perhaps worse than the 1970s rate of decline. If anything, employment decentralization has fared somewhat worse than population decentralization using this measure (table 2). Population loss of share has improved over time; the pace of

employment share loss has deteriorated or, at least, remained about the same. Perhaps the inner suburbs of midwestern metropolitan areas are also faring poorly as job locales in relation to the periphery. At least it appears that they are not doing well enough to pull up measured central county employment in relation to the peripheral counties of the metropolitan regions. Employment share erosion of the suburban portion of the central county is consistent with the findings of Myron Orfield, who documents that the problems once thought to characterize large inner cities—loss of tax base, population, and jobs—are now typical of the inner ring suburbs of older "inelastic" cities as well.[8]

Can we corroborate the finding of city job site decline any further? Comprehensive data on jobs by location over time are extremely spotty at the city level of geography—at least with regard to data sets that are consistently constructed so as to be comparable from state to state. However, I can use data from the censuses of business to shed some light on city-specific employment trends in the 1990s versus earlier decades. The business censuses do report accurately payroll employment by city geography. The downside is that coverage of industries is incomplete. Several service sectors are not covered for years before 1987, along with finance, insurance, real estate, transportation, communication, and public utilities. These are admittedly some sizable industries, and some of those that we know from other data sources to be most prominent (and central city durable) in central city locales. Nonetheless, a sizable amalgam of total employment remains that can be used to construct a "total employment" measure, comprising manufacturing, retail trade, wholesale trade, services (part), and government (part). The Census Bureau estimates that the business census data cover 75 percent of total payroll employment for 1987.[9]

We can see that the data trends displayed for central counties tend to be confirmed—even magnified—according to the business census data. On the whole for the 11 cities, the decline in the average annual employment trend accelerated from 1977–87 to 1987–97 (table 9). In measuring each city as an observation with equal weight, employment growth from 1977 to 1987 turned from slightly positive on an average annual basis to a negative annual decline of 1.4 percent per year during the 1987–97 period. This pattern was repeated for the city performance taken in aggregate—the so-called weighted average. Here, Chicago's large size and somewhat superior performance pulls up the average for all 11 cities. It is also notable that these city job losses were a stark contrast to the pace of job growth in the overall MSAs, which experienced gains of over 1 percent per year over the latter period. The consequences of

TABLE 8

Average annual change in central county employment

	County employment growth			Share of MSA employment		
	1969–79	1979–89	1989–98	1969–79	1979–89	1989–98
	(percent)			*(percentage points)*		
Buffalo	0.4	0.8	0.5	0.0	0.1	0.1
Chicago	0.3	0.6	0.7	–0.8	–0.6	–0.9
Cincinnati	1.1	1.3	1.1	–0.7	–0.6	–1.1
Cleveland	0.0	0.2	0.8	–0.7	–0.3	–0.6
Columbus	2.6	3.1	2.4	–0.1	0.3	–0.2
Detroit	–0.8	–1.0	–0.3	–1.9	–2.1	–1.7
Indianapolis	1.2	1.8	2.1	–0.5	–0.3	–0.5
Milwaukee	1.1	0.3	0.2	–0.9	–0.7	–1.3
Minneapolis-St. Paul	2.2	2.1	1.6	–0.8	–0.6	–0.7
Pittsburgh	0.5	0.4	0.8	–0.3	0.3	–0.4
St. Louis	3.2	3.7	1.3	1.8	2.1	0.1
All 11 central counties (weighted avg.)	0.7	0.9	0.9	–0.6	–0.4	–0.7
All 11 central counties (unweighted avg.)	1.1	1.2	1.0	–0.5	–0.2	–0.6
All 11 MSAs	1.4	1.3	1.6	n.a.	n.a.	n.a.
U.S.	2.3	2.1	1.8	n.a.	n.a.	n.a.

Notes: n.a. indicates not applicable. MSA is metropolitan statistical area.
Source: U.S. Department of Commerce, Bureau of Economic Analysis, various years, *Regional Economic Information System.*

this city–suburban disparity are that the central city lost share to the suburbs in the second period, and did so at an accelerated rate of 2 percent to 2.5 percent loss of share per year in 1987–97 versus the pace of 1 percent to 1.5 percent per year in the 1977–87 period. The generally buoyant Midwest economy has not lifted the central city as job domicile over the recent period in relation to the suburbs, though some central cities, such as Chicago, have bucked the trend. There has not been any slowing in the pace of erosion of job share for the central city. In observing this subset of payroll jobs, the evaporation of the city's importance in the wide metropolitan area seems to be accelerating.

TABLE 9

Annual average change in city employment
(percent)

	City employment		City share of MSA employment	
	1977–87	1987–97	1977–87	1987–97
Buffalo	–0.4	–2.9	–0.7	–3.4
Chicago	–1.3	0.5	–2.3	–0.4
Cincinnati	1.3	–2.0	–1.0	–3.3
Cleveland	–1.9	–1.9	–2.5	–2.6
Columbus	3.5	1.7	0.0	–0.4
Detroit	–2.1	–3.6	–3.0	–4.1
Indianapolis	2.5	1.3	0.2	–0.8
Milwaukee	–0.6	–0.8	–1.5	–2.2
Minneapolis-St. Paul	2.3	–2.8	–1.4	–4.2
Pittsburgh	0.4	–1.9	0.6	–2.8
St. Louis	–0.1	–3.2	–1.7	–3.8
Weighted avg.	–0.2	–1.0	–1.5	–2.0
Unweighted average	0.3	–1.4	–1.2	–2.5
Weighted average of 11 MSAs	1.5	1.3	n.a.	n.a.
U.S.	3.0	3.8	n.a.	n.a.

Notes: n.a. indicates not applicable. MSA is metropolitan statistical area. Total employment as calculated from business census data for manufacturing, wholesale, retail, services, and government for 1977, 1987, and 1997. Government employment reflects only local government employment for the MSAs and the U.S. and only municipal employment for the city.
Source: Business census data for manufacturing, wholesale, retail, services, and government for 1977, 1987, and 1997.

Conclusion

The central cities of the Midwest's large metropolitan areas are riding the favorable growth trends of the overlying Midwest economy. The 1970s were a terrible decade for central cities that followed upon the tumultuous times of the 1960s. Despite a profound Midwest recession that unfolded during the first three years of the 1980s, subsequent economic recovery was strong enough to make the 1980s look like an improvement over the 1970s. The late 1980s and 1990s solidified and magnified overall gains in the Midwest economy. As a consequence, central cities are now enjoying very strong rates of work force participation, a slowing of population loss, and rising real household incomes. Nonetheless, when we look beneath these statistics for signs of a structural change that would indicate that cities may regain their former prominence, there is less to cheer about. Relative to their suburbs, and accounting for the national business cycle, cities are faring little better than the 1980s (though better than the 1970s along some dimensions). Average household income in central cities relative to their suburbs continues to erode. Central city residents are finding employment, but increasingly in the suburbs. As the domicile of job location, central cities appear less attractive in the 1990s in relation to their suburbs, at least in terms of the pace of loss of share.

Of course, there may be evidence of revival that underlies these broad and aggregate statistics. So too, there are exceptional cities that are flashing recovery statistics, such as Chicago, that may be studied for clues to success and redevelopment. And the bright side should not be discounted. The improved absolute conditions brought about by U.S. economic expansion and Midwest revival in the 1990s may provide the foundation and resources on which to fashion an urban revival. However, this look at the current trends for improvement in the structural growth of central cities does not justify any complacency on the part of urban leaders and policymakers.

NOTES

[1]The circumstances of annexation differ greatly from city to city. The state legislature mandated a merger between the old city of Indianapolis and most of its surrounding county area. Indianapolis then merged many of its services with the remainder of Marion County as of 1970 into what is called Unigov. However, schools remain part of independent local governments, and townships remain, which include fire and relief responsibilities. So too, police services remain part of the former city of Indianapolis, while four former suburbs were allowed to retain their independence. In Columbus, Ohio, a forward-looking mayor named Jack Sensenbrenner adopted an aggressive policy of trading municipal services for annexation in the 1950s, allowing that city to gather up prime land around the emerging interstate highways and beltways in the 1960s and beyond. Milwaukee used its monopoly over Lake Michigan water to encourage annexation in the post WWII era. Milwaukee mayors were mostly annexation-minded throughout the first half of the century, though the city met resistance from industrial intensive fringe areas that feared higher property tax rates. A state legislative statute largely greatly impeded city annexation in 1956 by greatly easing the ability of mostly rural areas surrounding Milwaukee to incorporate.

The reasons some cities vigorously annexed and others chose not to remain cloudy. Surely, some city leaders pursued a self-interested fiscal calculus in pursuing annexation. For example, see Saffran (1952). Dye's (1964) study of U.S. urbanized areas for 1960 concluded that age of central city, social inequity between city and suburb, and form of government were partially explanatory factors. For a review of related studies, see Klaff and Fuguitt (1978).

[2]For a discussion see Mieszkowski and Mills (1993) and Brueckner (2000).

[3]See White (1999).

[4]Models have been explored in which capital stock, once built, is either abandoned or remains forever, or is durable but replaceable. So too, initial investment may take place myopically, or with degrees of foresight. See Wheaton (1983).

[5]Regression analysis confirms this finding; Indianapolis may be an exception in that average city household incomes appear to have strengthened in the 1990s.

[6]See U.S. Department of Commerce, Bureau of the Census, (various years), *Journey to Work* statistics.

[7]These data gather county level statistics from a number of sources so as to achieve complete industry coverage; estimates of self-employed workers along with payroll workers are included in the data.

[8]Orfield and Rusk (1998).

[9]Economic census data covered 75 percent of "economic activity" in 1987. In 1992, it covered 98 percent. See Micarelli (1998), p. 372.

REFERENCES

Aaronson Daniel, and Daniel G. Sullivan, 1998, "Recent trends in job displacement," *Chicago Fed Letter*, December, No. 136.

Brennan, John, and Edward W. Hill, 1999, *Where are the Jobs? Cities, Suburbs, and the Competition for Employment*, Washington DC: The Brookings Institution, Center on Urban and Metropolitan Policy, November.

Brueckner, Jan K., 2000, "Urban sprawl: Diagnosis and remedies," *International Regional Science Review*, Vol. 23, No. 2, April, pp. 160–171.

_____, 1987, "The structure of urban equilibria: A unified treatment of the Muth–Mills model," in *Handbook of Urban Economics, Vol. II*, Edwin S. Mills (ed.), Amsterdam: Elsevier Science B.V.

Dye, T. R., 1964, "Urban political annexation: Conditions associated with annexation in American cities," *Midwest Journal of Political Science*, Vol. 8, No. 4, pp. 430–446.

Gurda, John, 1999, *The Making of Milwaukee*, Milwaukee, WI: Milwaukee County Historical Society.

Illinois Department of Employment Security, 2000, *Where Workers Work, 1999*, Chicago, IL.

Klaff, Vivian Z., and Glenn V. Fuguitt, 1978, "Annexation as a factor in the growth of cities, 1950–60 and 1960–70," *Demography*, Vol. 15, No. 1, February, pp. 1–12.

Knepper, George W., 1997, *Ohio and Its People*, Kent, OH: Kent State University Press.

McMillen, Daniel P., and John F. McDonald, 1998, "Suburban subcenters and employment density," *Journal of Urban Economics*, Vol. 43, No. 2, March, pp. 157–186.

Micarelli, William F., 1998, "Evolution of the United States economic censuses: The nineteenth and twentieth centuries," *Government Information Quarterly*, Vol. 15, No. 3.

Mieszkowski, Peter, and Edwin S. Mills, 1993, "The causes of suburbanization," *Journal of Economic Perspectives*, Vol. 7, No. 3, Summer, pp. 134–147.

Mills, Edwin S., 1972, *Studies in the Structure of the Urban Economy*, *Resources for the Future*, Baltimore, MD: Johns Hopkins Press.

Mills, Edwin S., and Luan Sende Lubuele, 1997, "Inner cities," *Journal of Economic Literature*, Vol. 35, June, pp. 727–756.

Orfield, Myron W., and David Rusk, 1998, *Metropolitics: A Regional Agenda for Community and Stability,* Washington, DC: The Brookings Institution, June.

Pagano, Michael A., 1999, "Metropolitan limits: Intrametropolitan disparities and governance in U.S. laboratories of democracy," in *Governance and Opportunity in Metropolitan America*, Alan Altshuler, William Morrill, Harold Wolman, and Faith Mitchell (eds.), Washington DC: National Academy Press, pp. 253–295.

Saffran, George C., 1952, *Annexation Practices in Milwaukee, Study 52034*, Milwaukee, WI: City of Milwaukee, Budget Supervisor, Administration Survey Committee, June.

U.S. Department of Commerce, Bureau of Economic Analysis, 1998, *Regional Economic Information System*, Washington, DC.

U.S. Department of Commerce, Bureau of the Census, 1997, *Census of Manufactures*, Washington, DC.

_____, 1992, *Census of Manufactures*, Washington, DC.

_____, 1987, *Census of Manufactures*, Washington, DC.

_____, 1982, *Census of Manufactures*, Washington, DC.

_____, 1977, *Census of Manufactures*, Washington, DC.

_____, various issues, *Census of Population*, Washington, DC.

_____, 1997, *Census of Retail Trade*, Washington, DC.

_____, 1992, *Census of Retail Trade*, Washington, DC.

_____, 1987, *Census of Retail Trade*, Washington, DC.

_____, 1982, *Census of Retail Trade*, Washington, DC.

_____, 1977, *Census of Retail Trade*, Washington, DC.

_____, 1997, *Census of Selected Services*, Washington, DC.

_____, 1992, *Census of Selected Services*, Washington, DC.

_____, 1987, *Census of Selected Services*, Washington, DC.

_____, 1982, *Census of Selected Services*, Washington, DC.

_____, 1977, *Census of Selected Services*, Washington, DC.

_____, 1997, *Census of Wholesale Trade*, Washington, DC.

_____, 1992, *Census of Wholesale Trade*, Washington, DC.

_____, 1987, *Census of Wholesale Trade*, Washington, DC.

_____, 1982, *Census of Wholesale Trade*, Washington, DC.

_____, 1977, *Census of Wholesale Trade*, Washington, DC.

_____, various issues, *County Business Patterns,* available on the Internet at www.census.gov/epcd/cbp/view/cbpview.html.

U.S. Department of Commerce, Bureau of the Census, and U.S. Department of Labor, Bureau of Labor Statistics, various years, *Current Population Survey*, available on the Internet at www.bls.census.gov/cps/cpsmain.htm.

U.S. Department of Housing and Urban Development, 2000, *The State of the Cities 2000*, Washington, DC.

Wheaton, William C., 1983, "Theories of urban growth and metropolitan spatial development," in *Research in Urban Economics: A Research Annual, Vol. 3*, J. Vernon Henderson (ed.), Greenwich, CT: JAI Press Inc., pp. 3–36.

White, Michelle J., 1999, "Urban areas with decentralized employment: Theory and empirical work," in *Handbook of Regional and Urban Economics*, E. S. Mills and P. Cheshire (eds.), Amsterdam: Elsevier Science B.V.

Polycentric urban structure: The case of Milwaukee

Daniel P. McMillen

Introduction and summary

Theoretical models of urban structure are based on the assumption that all jobs are located in the central business district (CBD). Although this assumption was never literally true, it is a useful approximation for a traditional city in which the CBD holds the only large concentration of jobs. As metropolitan areas have become increasingly decentralized, traditional CBDs have come to account for a much smaller proportion of jobs than in the past. Large employment districts have arisen outside of central cities that rival the traditional city center as places of work. When these districts are large enough to have significant effects on urban spatial structure, they are referred to in the urban economics literature as "employment subcenters."

The distinction between a metropolitan area with multiple subcenters (or a polycentric urban structure) and one with much more dispersed suburban employment has important policy implications. Public transportation can be designed to serve subcenters. Buses can help alleviate severe congestion, and commuter rail lines may be able to serve large subcenters. Large subcenters may have enough jobs to warrant designing public transportation that brings central city workers to suburban job locations, which can help alleviate problems of a "spatial mismatch" between jobs and central city workers (Kain, 1968, and Ihlanfeldt and Sjoquist, 1990). The term "urban sprawl" appears to be used to describe an urban area whose residents have moved farther and farther from the central city, while driving past pockets of farmland and open space to get to their suburban jobs. Sprawl is likely to be less of a problem in an urban area whose suburban jobs are concentrated in subcenters. If jobs are confined to a relatively small number of suburban sites, workers will attempt to reduce their commuting costs by living nearby. This tendency toward suburban centralization is reinforced when transportation facilities are designed to serve the subcenters.

Spatial modeling of traditional monocentric cities is relatively easy because the site of the CBD is known in advance. Housing prices, land values, population density, and other variables of interest can be modeled as functions of distance to the CBD, with the addition of other variables of local importance, such as distance to Lake Michigan in Chicago or proximity to freeway interchanges and commuter train stations. In contrast, subcenter locations are not always obvious beforehand. The U.S. Census lists central places, which are generally older suburbs that once were satellite cities. However, subcenters are often relatively new developments (dubbed "edge cities" by Garreau, 1991) that may not have been incorporated as recently as 1960. Subcenter locations are an empirical issue: Does an area have enough employment that it has a significant local effect on variables such as employment density?

In this article, I critique various procedures for identifying employment subcenters and then use a procedure developed in McMillen (2002) to analyze subcenters in Milwaukee, Wisconsin. Milwaukee is interesting because it has not been the subject of a great deal of study, yet it is representative of older industrial cities that have maintained strong CBDs. I identify subcenters as local peaks in an estimated employment density function. I find that Milwaukee has one subcenter, which is located at the western edge of the city. It is notable for being the site of a Harley-Davidson manufacturing plant, although other

Daniel P. McMillen is a professor of economics at the University of Illinois at Chicago and a consultant to the Federal Reserve Bank of Chicago.

firms also are located in the area. The subcenter has significant but highly localized effects on both employment and population densities in the Milwaukee area. Milwaukee remains a largely monocentric city.

Although Milwaukee has a monocentric spatial structure, it has ample suburban employment that is highly dispersed. Its single subcenter is readily accessible by central city residents, but the subcenter has fewer than 25,000 jobs in a metropolitan area of 821,158 workers. The dispersed nature of Milwaukee's suburban jobs makes it difficult to design a public transportation system that would help carry central city residents to suburban jobs. Milwaukee's dispersed employment increases the probability of central city unemployment and increases urban sprawl as suburban residents move still farther from the central city.

The rise of the polycentric city

The monocentric city model of Alonso (1964), Muth (1969), and Mills (1972) remains the most popular and influential model of urban spatial structure. The model depicts a stylized nineteenth century city, in which all jobs are located in the CBD. To reduce the cost of their daily commute, workers bid more for housing close to the city center. As a result, housing and land prices are predicted to fall with distance from the CBD. Spatial patterns for other variables of interest—population density, lot sizes, building heights, and the like—are all predicted to be simple functions of distance from the CBD.

Although these predictions have ample empirical support,[1] the central idea of the monocentric city model—that urban employment is concentrated in the traditional CBD—is no longer a suitable representation of urban spatial structure. Indeed, McDonald and McMillen's (1990) evidence of multiple peaks in land value functions in early twentieth-century Chicago suggests that the assumption of monocentricity was always more of a mathematical convenience than an accurate depiction of reality. Recent theoretical and empirical research in urban economics treats metropolitan areas as polycentric, that is, having multiple employment centers with varying degrees of influence on urban spatial patterns. Anas, Arnott, and Small (1998) present an excellent survey of theoretical and empirical models of polycentric cities.

The polycentric structure of urban areas has become more evident over time. Table 1 presents evidence of declining employment concentration in 11 midwestern urban areas. Across all 11 cities, 36.6 percent of suburban residents worked in the central city in 1960, whereas only 9.4 percent of city residents worked in the suburbs. The percentage of suburban residents working in the city ranged from 16.8 percent in Pittsburgh to 62.3 percent in Indianapolis. By 1990, the percentage of suburban residents working in the city had declined in every metropolitan area except Pittsburgh. Overall, only 28.4 percent of suburban residents worked in the central city in 1990, while 26.2 percent of city residents worked in the suburbs. Pittsburgh is an outlier because the large

TABLE 1

Journey to work patterns

	City residents working in the suburbs				Suburban residents working in the city			
	1960	1970	1980	1990	1960	1970	1980	1990
Buffalo	17.1	26.8	25.3	27.9	36.5	30.4	28.0	29.6
Chicago	6.6	16.1	18.4	22.5	34.6	27.1	22.5	25.6
Cincinnati	11.2	24.8	24.4	29.5	45.0	39.5	36.3	31.6
Cleveland	7.7	24.4	28.6	30.3	52.4	43.5	34.8	32.5
Columbus	7.8	19.1	17.7	24.2	50.6	54.5	48.1	49.7
Detroit	17.3	32.1	34.3	36.4	33.5	24.6	16.9	19.4
Indianapolis	6.1	18.5	9.8	12.1	62.3	44.8	48.7	48.9
Milwaukee	8.9	23.7	26.3	30.3	48.0	36.1	33.7	37.2
Minneapolis-St. Paul	6.6	19.7	24.5	29.8	52.1	43.5	31.2	30.5
Pittsburgh	11.2	19.1	20.1	21.4	16.8	24.6	26.4	24.3
St. Louis	8.3	21.1	24.0	35.9	36.7	30.0	25.4	27.9
All	9.4	21.2	21.8	26.2	36.6	31.8	27.0	28.4

Note: Data for 1990 reflect all central cities in the consolidated metropolitan statistical areas.
Source: U.S. Department of Commerce, Bureau of the Census, various years.

suburban steel plants closed during this period, leading to renewed employment centralization. Table 1 clearly shows that the CBD is not the dominant employment site in any of these cities, and that city residents are now nearly as likely to work in the suburbs as suburban residents are to work in the city.

The diminishing role of the CBD has come about despite the advantages it offers for firms wishing to locate in metropolitan areas. In-place public transportation, such as light rail, and radial boulevards and highways are designed to carry workers from outlying areas into the city. Reverse commuting and intrasuburban commuting is very difficult other than by automobile. Whereas highways lead from many directions in to the city, a suburban firm may find that its potential labor pool is limited to a relatively small geographic area around the workplace. In addition, theories of agglomeration such as Anas and Kim (1996), Berliant and Konishi (2000), and Fujita and Ogawa (1982) suggest that firms may enjoy significant cost advantages by locating near other firms. The close proximity of firms in the CBD facilitates face-to-face communication. Lawyers, bankers, and myriad consultants are all nearby in the CBD. Both suppliers and customers are likely to require only a short trip to visit a CBD firm.

But suburban locations offer different advantages. Land is significantly cheaper than in the CBD, and access to interstate highways is better and subject to less congestion. Large manufacturing firms are more likely to prefer suburban locations, as are distributors and wholesalers that have customers outside the metropolitan area. Suburban locations may reduce the wage bills of firms whose workers live in the suburbs because less compensation is needed for an expensive and time-consuming commute.

Employment subcenters combine many of the advantages of CBD and suburban locations. Highways and public transportation can serve subcenters much as they serve the CBD, bringing in an ample supply of workers from distant locations. Costs may be lower than in the CBD because land is cheaper and many workers like to live and work in the suburbs. Personal communication may be as easy as in the CBD when firms locate near one another in subcenters. Restaurants and other services find enough business to form concentrations in the vicinity. The diversity of business types may be lower than in the city, but large subcenters sometimes appear to mimic the diversity of CBDs while offering lower land and commuting costs. Large subcenters offer employment and shopping opportunities for which nearby residents are willing to pay a premium. As predicted by the monocentric city model for locations near the CBD, the rise in land values near subcenters leads to configurations with smaller lot sizes and higher population density that look like small cities.

Subcenter identification procedures

Empirical researchers have long recognized that cities are not truly monocentric. Variables representing distance from various employment sites other than the CBD are frequently included as explanatory variables in empirical studies of housing prices, employment density, and population density.[2] Sites that are significant enough to affect the overall urban spatial structure must be specified beforehand using this ad hoc approach. Forming the list of potential subcenters often draws on ample local knowledge, but may well be inconsistent with the data. Although statistically insignificant subcenter distance variables help indicate that the subcenter list is incorrect, they do not reveal subcenter sites that are omitted from the regressions.

The first formal procedure for identifying employment subcenters was proposed by McDonald (1987). He begins by estimating a simple employment density function for a standard monocentric city: $log\ y_i = \alpha + \beta x_i + \varepsilon_i$, where y_i represents the number of employees per acre and x_i is distance from the CBD. Subcenters produce clusters of positive residuals in the estimated function. McDonald inspects the list of statistically significant positive residuals, and finds that O'Hare Airport is the dominant subcenter in the Chicago metropolitan area.

McDonald's novel approach poses several problems in practice. The notion of a "cluster" is subject to interpretation. Are two significant positive residuals among ten observations in a two-mile radius a cluster? A reasonable change in either the radius or the requisite number of positive residuals can potentially change the results dramatically. The procedure also suffers from statistical problems. The results are sensitive to the unit of analysis. Using extremely large tracts, McDonald (1987) finds a single subcenter in the Chicago area near O'Hare Airport. In a follow-up paper using square mile tracts, McDonald and Prather (1994) find additional subcenters in Schaumburg and central DuPage County. The local rise in employment density produced by a subcenter tends to flatten the estimated employment density function, which reduces the probability of identifying subcenters. Although the monocentric employment density function implies that gradients do not vary across the urban area, multiple subcenters or distinctive topographical features may lead to variations in gradients. Such functional form misspecification can hide potential subcenters.

Giuliano and Small (1991) propose another influential subcenter identification procedure. It has been employed in subsequent work by Bogart and Hwang (1999), Cervero and Wu (1997, 1998), and Small and Song (1994). Defining a subcenter as a set of contiguous tracts that have a minimum employment density of 10 employees per acre each and, together, have at least 10,000 employees, Giuliano and Small identify 32 subcenters in the Los Angeles area. This reasonable subcenter definition is sensitive to the cutoff points used for minimum employment density and total subcenter employment. The same cutoff points imply an unreasonably large subcenter in the northern Chicago suburbs with over 400,000 employees, leading McMillen and McDonald (1998) to raise the cutoffs to 20 employees per acre and 20,000 total employees. Local knowledge must guide the choice of cutoff points, limiting the analysis to familiar metropolitan areas.

Giuliano and Small's procedure is also sensitive to the unit of analysis. Their data set includes 1,146 tracts covering an area of 3,536 miles. In contrast, McMillen and McDonald's Chicago data set has 14,290 tracts in an area of 3,572 square miles. Data sets with small tracts are more likely to have pockets with low employment density, which reduces the number of subcenters identified using the Giuliano and Small procedure. This observation led McMillen and McDonald (1998) to work with proximity instead of contiguity: Two tracts are proximate to one another if they are within 1.5 miles. The number of subcenters is again sensitive to the definition of proximity.

Giuliano and Small define a subcenter as an area with large employment, with the definition of "large"—the cutoff points—being up to the analyst. Subsequent statistical analysis determines whether the subcenters have significant effects on such variables as employment density, population density, and housing prices. The cutoffs do not vary over the data set, which means that the minimum subcenter size is the same near the CBD as in distant suburbs. This characteristic of their procedure is not desirable if a subcenter is defined as an area with larger employment density than surrounding areas. Since densities tend to decrease with distance to the CBD, the minimum cutoffs should tend to decrease also. Then the question becomes how to vary the cutoffs.

Craig and Ng (2001) propose a procedure that eliminates many of the problems with the earlier methods. They use a nonparametric estimation procedure to obtain smoothed employment density estimates for Houston. Using a quantile regression approach, they focus on the 95th percentile of the employment density distribution. The quantile regression approach is attractive in this context because a subcenter is defined using the extremes of the distribution. Craig and Ng's estimated density function is symmetric about the CBD because they only use distance from the CBD as an explanatory variable for the estimates. They first look for local rises in the density–CBD relationship, and then inspect the rings to find sites with unusually high density and employment. They use their knowledge of Houston to accept or reject high-density sites as subcenters.

Craig and Ng's procedure is not as sensitive to the unit of analysis as the McDonald and Giuliano–Small procedures. Though larger tracts lead to smoother employment density functions, a large subcenter will produce a rise in the function whether the data set includes acres, quarter sections, or square miles. The procedure is readily reproducible by other researchers and requires scant knowledge of the local area. Much of the arbitrariness of the Giuliano–Small procedure is eliminated because the local rise that defines a subcenter is subject to tests of statistical significance. However, the Craig–Ng procedure requires some local knowledge to choose which sites are subcenters within rings around the CBD, and the imposition of symmetry around the CBD is unsuited to cities that are distinctly asymmetric due to varied terrain or multiple subcenters.

A nonparametric subcenter identification procedure

Nonparametric approaches offer significant advantages over simple linear regression procedures. Nonparametric estimators are flexible, allowing the slope of density functions to vary across the metropolitan area. As an example, suppose that employment density declines more rapidly on the north side of the city than on the south. The standard linear regression estimator used by McDonald (1987) imposes the same gradient on both sides of the city, which tends to produce positive residuals on the north side and negative residuals to the south. This functional form misspecification increases the probability of finding a subcenter on the north side of the city even if none exists. Craig and Ng's (2001) estimator is more flexible than standard linear regression, but does not avoid this type of misspecification because it imposes symmetry about the CBD. In contrast, nonparametric estimation procedures are sufficiently flexible to detect the difference in gradients across the two sides of the city.

McMillen (2002) proposes a nonparametric procedure for identifying subcenters in a variety of cities,

including those with which the analyst is largely unfamiliar. It is a two-stage procedure that combines features of both the McDonald (1987) and Craig and Ng (2001) approaches. As in McDonald (1987), the first stage of the procedure identifies subcenter candidates through an analysis of the residuals of a smoothed employment density function. The procedure differs in that McMillen uses a nonparametric estimator, locally weighted regression, to estimate the employment density function.[3] The estimation procedure involves multiple applications of locally weighted regression. McMillen estimates a separate regression for locations for which a log-employment density estimate is desired. Observations closer to the target location receive more weight in the regressions. McMillen (2002) identifies subcenter candidates as significant residuals (at the 5 percent level) from the first-stage locally weighted log-density estimates. When significant residuals cluster together, he narrows the list of subcenter candidate sites to those with the highest predicted log-employment density among all observations with significant positive residuals in a three-mile radius.

The second stage of the procedure uses a *semiparametric* procedure (Robinson, 1988) to assess the significance of the potential subcenter sites in explaining employment density. The nonparametric part of the regression controls in a general way for the nuisance variable, *DCBD*, which is an acronym for *distance from the central business district*. Following Gallant (1981, 1983, and 1987), McMillen (2002) uses a flexible Fourier form to approximate the nonparametric part of the regression (see box 1). Distances to potential subcenter sites are included as explanatory variables in the parametric part of the regression. If the regression indicates that densities fall significantly with distance from a potential subcenter site, then the site is included in the final list of subcenters.

This procedure reflects the definition of subcenters listed earlier: Subcenters are sites that cause a significant local rise in log-employment densities, after controlling for distance from the CBD. Unlike Giuliano and Small (1991), McMillen (2002) uses statistical tests to determine the significance of subcenter sites. This feature makes it possible to apply the procedure for a variety of cities, including unfamiliar ones. Basing the procedure on a semiparametric regression analysis allows the analyst to conduct statistical tests of significance, while reducing the sensitivity of the analysis to restrictive functional form specifications, the size of the unit of observation, and the specification of arbitrary cutoff points.

Data

The data come from the Urban Element of the Census Transportation Planning Package, which is produced by the Department of Transportation's Bureau of Transportation Statistics (BTS). The BTS produced special tabulations of 1990 U.S. census data to match standard census data with their unit of analysis, which they term the *transportation analysis zone*, or "taz." The zones vary in size across metropolitan areas, but are usually smaller than census tracts or zip codes. All data for this study cover the Milwaukee metropolitan area, which comprises Milwaukee, Kenosha, Ozaukee, Racine, Washington, and Waukesha counties.[4]

The taz sizes average 2.1 square miles in this sample of 1,206 observations. Total population is 1,805,245, and total employment is 821,158, or 45.5 percent of the population. Average densities imply that population is more dispersed than employment. Employment density averages 2,598 workers per square mile, or 4.1 employees per acre. In contrast, population density averages 3,244 people per square mile, or 5.1 people per acre.

The Milwaukee subcenter

Figure 1 presents a map showing employment densities in the Milwaukee area. Aside from pockets of high densities in Racine and Kenosha, the map suggests that Milwaukee is not far from a stylized monocentric city. This finding is reflected in the McMillen (2002) procedure, which identifies a single employment

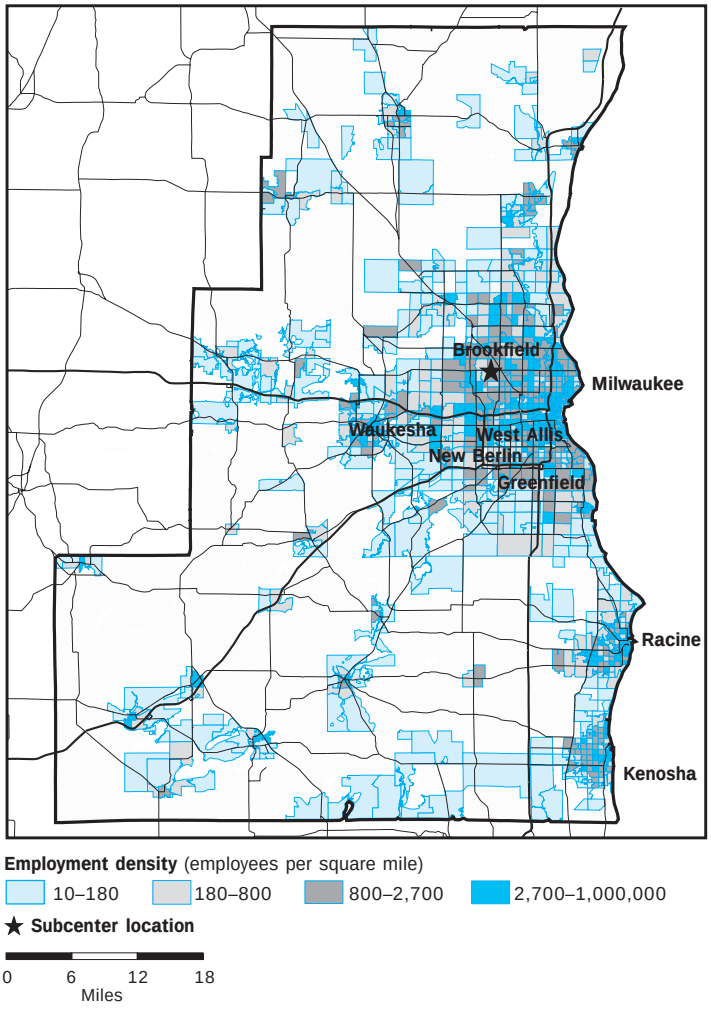
BOX 1

Fourier terms

The Fourier expansion uses sine and cosine terms to approximate the general function $g(DCBD)$. To implement the procedure, the variable *DCBD* is first transformed to lie between 0 and 2π, with the transformed variable denoted by z. The Fourier expansion is $g(DCBD_i) \approx \lambda_0 + \lambda_1 z_i + \lambda_2 z_i^2 + \Sigma_q(\gamma_q \cos(qz_i) + \delta_q \sin(qz_i))$, where $q = 1, \ldots, Q$. The Schwarz (1978) information criterion is used to choose the expansion length, Q. The optimal Q is the value that minimizes $S(m) = \log(s^2) + m\log(n)/n$, where m is the number of estimated coefficients ($m = 3 + 2Q$), s^2 is the estimated variance of the errors from the semiparametric regression, and n is the number of observations. Larger values of Q reduce the estimated variance but increase the second term. The subcenter distance variables are omitted when choosing Q.

FIGURE 1

Employment density in Milwaukee and subcenter location



Source: Author's calculations based on data from the U.S. Department of Commerce, Bureau of the Census, transportation planning package.

subcenter. Its location is shown in figure 1. The subcenter is at the edge of the City of Milwaukee, at the intersection of State Highway 45 and Route 190, near Wauwatosa. The site includes the main Harley-Davidson manufacturing plant. It meets the Giuliano and Small (1991) criterion for a subcenter by including two tracts with more than 10 employees per acre. The larger tract, which includes the Harley-Davidson plant, has 17.0 employees per acre and 10,344 total workers. The other tract has 10.5 employees per acre and 3,759 workers.

Table 2 provides more information on employment patterns in the Milwaukee area. The CBD is defined as an area one mile in diameter around the tract at the city center with the largest employment density. The subcenter is an area three miles in diameter around its midpoint. Both areas include 11 observations. Only 6.7 percent of Milwaukee's employment is in the CBD (as defined here), but the CBD is nonetheless more than twice as large as the subcenter, which has 3.0 percent of total employment in the metropolitan area. As predicted by urban theory, median earnings are highest in the CBD, but it is interesting to note that earnings on average are higher in the subcenter than in the rest of the city. The earnings differences are not large, but they suggest that either marginal productivity is higher in sites with high employment density or that firms must compensate workers for longer commutes. In keeping with the spatial mismatch hypothesis, African-Americans comprise a larger percentage of total employment in the CBD. In contrast to the spatial mismatch hypothesis, however, this tendency toward CBD employment may *increase* the average earnings of African-Americans because average earnings are lower elsewhere. In part because the subcenter is only 8.1 miles from the CBD, the percentage of African-Americans in the subcenter is closer to that in the CBD than in the rest of the city. This result is significant because it indicates that the commute to a nearby subcenter may be only slightly more burdensome than a commute to the CBD for central city residents.

Table 2 shows the employment mix in the CBD, subcenter, and the rest of the city for five traditional industry categories. The CBD specializes in the financial, insurance, and real estate sector (26.61 percent of CBD employment) and service industries (34.27 percent of CBD employment). In contrast, a larger percentage of the subcenter's employment (30.48 percent) is engaged in manufacturing, with a significant concentration in retail also. Service industries are underrepresented in the subcenter compared with the CBD or the rest of the city. On the whole, the employment mix in the subcenter is closer to the mix in the rest of the city than to the CBD.

Comparison of employment density estimates

Figure 2 presents graphs of the estimated log-employment densities along a ray from the CBD to the subcenter. The grey line shows that the initial

TABLE 2

Employment mix

	CBD	Subcenter	Rest of city
Total employment	54,669	24,967	741,522
Number of residents	4,508	19,260	1,781,477
Median earnings ($)	21,397	20,715	19,064
	(- - - - - - - % of total employment- - - - - - -)		
White	87.06	89.29	89.60
Black	9.75	9.05	7.82
Manufacturing	10.92	30.48	26.13
Transportation, communications, utilities, and wholesale	11.08	10.85	10.57
Retail	8.79	23.48	17.03
Financial, insurance, and real estate	26.61	9.99	5.56
Services	34.27	21.95	31.91

Note: CBD is central business district.
Source: Author's calculations based on data from the U.S. Department of Commerce, Bureau of the Census, transportation planning package.

locally weighted regression estimates decline rapidly with distance from the CBD up to about 18 miles, after which the decline is nearly linear. The black line shows that the simple exponential function used by McDonald (1987) is badly misspecified here, indicating a much less rapid rate of decline in densities after about seven miles than found using the more flexible nonparametric estimator. The Fourier estimates detect a sharp rise in employment density around the subcenter, although they too tend to overestimate densities in distant locations. Figure 2 shows that McDonald's estimator would have trouble finding subcenters in distant areas because the overestimate of densities will tend to produce *negative* rather than positive residuals.

Just as simple exponential function overestimates densities along the ray between the CBD and the subcenter, figure 3 shows that it tends to underestimate densities along a ray due south from the CBD. Densities do not decline as rapidly on the south side of Milwaukee as to the north. Together, figures 2 and 3 show the advantages of locally weighted regression's flexibility over the symmetric McDonald (1987) and Craig–Ng (2001) estimators.[5] Figure 4 shows an advantage of the nonparametric approach over the Giuliano–Small (1991) procedure. The entire log-employment density function lies below the cutoff point of 10 employees per acre, which is why only two tracts—those with large positive residuals—meet the cutoff. If the cutoff were raised to 20 employees per acre, the Giuliano–Small procedure would miss the subcenter entirely. If the cutoff point were lowered too far, the subcenter would simply be part of the CBD, or it would be so large as to be meaningless (as found in McMillen and McDonald, 1998, for Chicago).

FIGURE 2

Estimated employment density functions—ray between CBD and subcenter

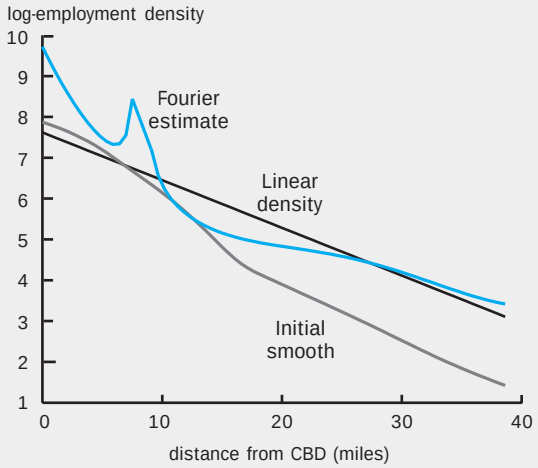


FIGURE 3

Estimated employment density functions—ray from CBD to south

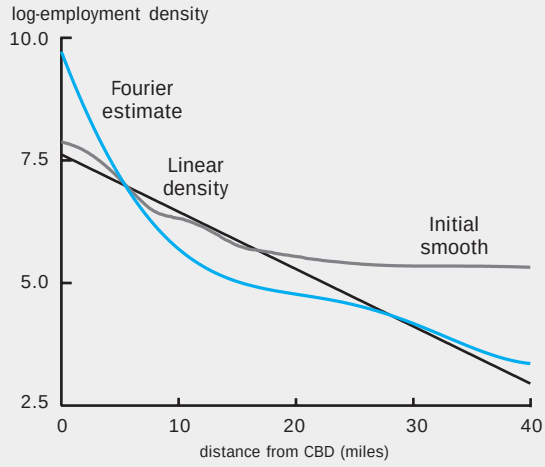
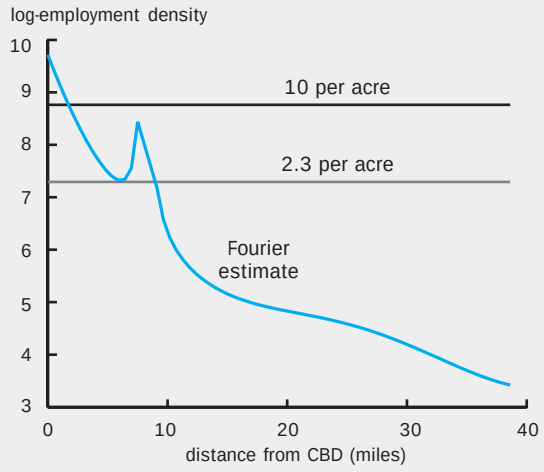


FIGURE 4

Cutoff points for Giuliano–Small method



Subcenters and urban sprawl

I define subcenters here as sites that cause significant *local* rises in employment densities. A question arises as to the extent of the subcenter's influence on the *overall* urban spatial structure. Traditionally in urban economics, urban decentralization is measured by the CBD gradient, which is the slope coefficient from a regression of the natural logarithm of population density on distance from the CBD (Clark, 1951; Macauley, 1985; McDonald, 1989; McDonald and Bowman, 1976; Mills, 1972; and Mills and Tan, 1980). The gradient measures the percentage decline in densities associated with a movement of one mile from the CBD. The relatively slow decline of densities in decentralized metropolitan areas is reflected in small gradients. Density gradients are thus a useful measure of urban sprawl.

The first column of results in table 3 presents the average gradients from various specifications of employment and population density functions. In a simple regression of log density on *DCBD*, employment density is estimated to decline by 11.7 percent and population density is estimated to decline by 7.6 percent with each mile from the CBD. These figures are consistent with those found previously for relatively centralized cities (for example, Macauley, 1985; or Mills and Tan, 1980). However, the apparent centralization of Milwaukee becomes more pronounced when more flexible functional forms are used in estimation. Flexible Fourier functions of *DCBD* imply much larger gradients: 28.2 percent per mile for employment density and 17.7 percent per mile for population density. Such steep declines in densities with distance to the CBD indicate a centralized urban area.

Milwaukee's subcenter has only a marginal impact on the estimated gradients. The gradients for distance from the CBD are virtually unchanged when the inverse of distance from the subcenter is added as an explanatory variable in the density regressions. For example, the employment density gradient only falls from –11.7 percent to –11.2 percent when the variable is added to a regression of log-employment density on *DCBD*. The second column of results in table 3 presents the corresponding gradients for distance from the subcenter, estimated using the same regressions as for the CBD gradients. The gradients, which are averages over the entire metropolitan area, are not statistically significant. Together, these results suggest that the subcenter has only a local effect on Milwaukee's spatial structure. It raises densities enough to have a statistically significant effect in the estimated functions, but not enough to be significant across the full metropolitan area or to cause severe bias in the estimated CBD gradients when omitted from the density functions.

The last column of table 3 presents the results of Lagrange multiplier (LM) tests for spatial autocorrelation (Anselin, 1988; Anselin et al., 1996; and Burridge, 1980). Spatial autocorrelation will be present if the residuals of the estimated density functions are correlated over space. If firms tend to cluster together, then the residuals of the employment density functions will be positively correlated spatially. The LM tests are thus a useful measure of spatial clustering. They are complementary to but different from our definition of a subcenter. Whereas a subcenter is an area with extremely high density, spatial autocorrelation may be found in areas without sharp peaks in density, yet with more clustering of employment than would be implied by random variation. Just as a metropolitan area with subcenters is less decentralized than an otherwise identical city with randomly distributed suburban employment, an area with a high degree of spatial autocorrelation in employment density is more centralized than an area with random variation in densities.

The LM tests presented in table 3 are highly significant in every case.[6] For the simple models in which only *DCBD* is included as an explanatory variable, the LM test statistics are 1,486.27 for employment density and 1,616.90 for population density. These values are far greater than the critical value of 3.84, and indicate an extremely high degree of spatial clustering of the residuals. The test statistics fall to 859.17 and 536.31 when the inverse of distance to the subcenter is added to the regressions. The decrease in the test statistics suggests that the residuals are

TABLE 3

Employment and population density

Explanatory variables	CBD gradient	Subcenter gradient	Spatial autocorrelation LM test
	Log-employment density		
Distance from CBD	–0.117 (0.006)		1,486.27
Fourier terms	–0.282 (0.073)		602.39
Distance from CBD and inverse of distance to subcenter	–0.112 (0.006)	–0.021 (0.020)	859.17
Fourier terms and inverse of distance to subcenter	–0.295 (0.074)	–0.033 (0.019)	592.13
	Log-population density		
Distance from CBD	–0.076 (0.004)		1,616.90
Fourier terms	–0.177 (0.037)		327.39
Distance from CBD and inverse of distance to subcenter	–0.074 (0.004)	–0.009 (0.009)	536.31
Fourier terms and inverse of distance to subcenter	–0.182 (0.037)	–0.013 (0.008)	321.20

Notes: The Fourier terms include z, z^2, $\cos(z)$, and $\sin(z)$, where z denotes the distance from the CBD multiplied by $2\pi/50$. See box 1, p. 19, for complete details on Fourier terms. Heteroscedasticity consistent standard errors (White, 1980) are in parentheses.

Source: Author's calculations based on data from the U.S. Department of Commerce, Bureau of the Census, transportation planning package.

much less clustered after allowing densities to rise near the subcenter. The higher degree of clustering in the model without the subcenter distance variable is a direct result of a large number of positive residuals near the subcenter site. Adding the Fourier expansion terms—z, z^2, $\cos(z)$, and $\sin(z)$—leads to further reductions in the LM test statistics. In the most general models, which include both the Fourier expansion terms and the inverse of distance to the subcenter, the LM test statistics are 592.13 for employment density and 321.20 for population density. Thus, the LM tests suggest that spatial autocorrelation remains significant even after controlling for the effects of the subcenter and when using a very general functional form for *DCBD*. Whereas estimated density functions imply that densities decline smoothly with distances from the CBD and subcenter, the spatial autocorrelation tests suggest that densities are in fact much more highly clustered than implied by smooth functions of distance.

Overall, these results indicate that Milwaukee remains a centralized city, although it has many suburban jobs. Even simple exponential functions imply large gradients for both employment and population density. More flexible functional forms imply still steeper gradients. Both employment and population are spread across Milwaukee in clusters, with densities that decline rapidly with distance from the city center.

Conclusion

Milwaukee's CBD still dominates metropolitan-wide employment and population density patterns. Nevertheless, jobs are spread throughout the metropolitan area. Table 1 shows that a majority of Milwaukee's suburban residents worked in the suburbs in 1990, and over 30 percent of its central city residents also worked in the suburbs. One area at the edge of the city is large enough to qualify for subcenter status. It is the location for a Harley-Davidson manufacturing plant and is the site for more than 20,000 jobs. The subcenter has significant effects on employment density and population density patterns in the vicinity. However, the effects are highly localized. Milwaukee is still primarily a monocentric city. Although it has ample suburban employment, the CBD dominates overall spatial density patterns in a manner largely consistent with Brueckner's (1979) version of the monocentric city model.

With only one subcenter set in the midst of ample suburban employment, little can be done in Milwaukee to relieve problems associated with congestion and a spatial mismatch between jobs and workers. If firms in the Milwaukee area had moved to a few large suburban subcenters, public transportation could be designed to carry commuters efficiently to suburban jobs. Central-city residents would not be at a serious disadvantage in taking suburban jobs if they could easily take buses to the large subcenters. Milwaukee's single subcenter can indeed be reached easily by central-city residents. However, the majority of Milwaukee's jobs are now scattered across the metropolitan area. This spatial pattern of employment opportunities makes it difficult for central-city residents to find jobs, and increases the probability that suburbanites will move still farther from the city center.

Researchers have identified subcenters for only a small number of cities—Chicago, Cleveland, Dallas, Houston, Los Angeles, New Orleans, the San

Francisco Bay Area, and now Milwaukee. It remains an open question whether there are systematic patterns across metropolitan areas concerning subcenters. Is there a critical population level at which subcenters become more likely? Are subcenters more likely in old or new cities or in cities with good public transportation service or those that rely predominantly on the automobile? Do subcenters increase the probability of reverse commuting and the probability of central city unemployment? Do subcenters increase the degree of sprawl by allowing suburbanites to live still farther from the center of the city? Do subcenters tend to specialize in particular types of employment, such as manufacturing or financial services? Recently developed procedures for identifying subcenters make it possible for researchers to answer these questions after determining the number, size, and employment mix of subcenters across metropolitan areas.

NOTES

[1]Examples include Clark (1951), Fales and Moses (1972), Macauley (1985), McDonald (1989), McDonald and Bowman (1976; 1979), McMillen (1996), and Mills (1969; 1970).

[2]Examples include Bender and Hwang (1985), Dowall and Treffeisen (1991), Gordon et al. (1986), Greene (1980), Griffith (1981), Heikkila et al. (1989), Richardson et al. (1990), and Shukla and Waddell (1991).

[3]Stone (1977) and Cleveland (1979) first proposed the locally weighted regression procedure, which has since been extended by Cleveland and Devlin (1988), Fan (1992, 1993), Fan and Gijbels (1992), and Ruppert and Wand (1994). It is a simple extension of the kernel regression estimator. Locally weighted regression has been used extensively in spatial modeling. Examples include Brunsdon et al. (1996), McMillen and McDonald (1997), McMillen (2002), Meese and Wallace (1991), Pavlov (2000), and Yuming and Somerville (2001).

[4]I used a mapping program to measure the area of each taz (in square miles) and to provide coordinates for the taz center points. These coordinates are used to measure distance to the CBD.

[5]As employed here, the Fourier estimator also imposes symmetry about the CBD. This misspecification is less critical in the second stage of the analysis, where the objective is only to assess the statistical significance of the subcenters. The misspecification could be eliminated by estimating $g(x_1,x_2)$ nonparametrically rather than $g(DCBD)$, where x_1 and x_2 represent distances north and east of the CBD.

[6]The test statistic is $(e'We/s^2)^2/\text{tr}(W'W + WW)$, where e is the vector of residuals and s^2 is the estimated variance of the regression. W is a "spatial contiguity matrix," representing the spatial relationship between observations. For the models in table 3, $W_{ij} = 1$ when observation j is among the nearest 1 percent of the observations to observation i, and $W_{ij} = 0$ otherwise. The rows of the $n \times n$ matrix W are then normalized such that each sums to one. The test statistic is distributed χ^2 with one degree of freedom, which implies a critical value of 3.84 for a test with a 5 percent significance level.

REFERENCES

Alonso, William, 1964, *Location and Land Use*, Cambridge, MA: Harvard University Press.

Anas, Alex, and Ikki Kim, 1996, "General equilibrium models of polycentric urban land use with endogenous congestion and job agglomeration," *Journal of Urban Economics*, Vol. 40, pp. 232–256.

Anas, Alex, Richard Arnott, and Kenneth A. Small, 1998, "Urban spatial structure," *Journal of Economic Literature*, Vol. 36, pp. 1426–1464.

Anselin, Luc, 1988, "Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity," *Geographical Analysis*, Vol. 20, pp. 1–17.

Anselin, Luc, Anil K. Bera, Raymond Florax, and Mann J. Yoon, 1996, "Simple diagnostic tests for spatial dependence," *Regional Science and Urban Economics*, Vol. 26, pp. 77–104.

Bender, B., and H. Hwang, 1985, "Hedonic housing price indices and secondary employment centers," *Journal of Urban Economics*, Vol. 17, pp. 90–107.

Berliant, Marcus, and Hideo Konishi, 2000, "The endogenous formation of a city: Population agglomeration and marketplaces in a location-specific production economy," *Regional Science and Urban Economics*, Vol. 30, pp. 289–324.

Bogart, William T., and H. Hwang, 1999, "Employment centers in Greater Cleveland: Evidence of evolution in a formerly monocentric city," *Urban Studies*, Vol. 36, pp. 2099–2110.

Brueckner, Jan K., 1979, "A model of non-central production in a monocentric city," *Journal of Urban Economics,* Vol. 6, pp. 444–463.

Brunsdon, Chris, A. Stewart Fotheringham, and Martin E. Charlton, 1996, "Geographically weighted regression," *Geographical Analysis,* Vol. 28, pp. 281–298.

Burridge, P., 1980, "On the Cliff–Ord test for spatial autocorrelation," *Journal of the Royal Statistical Society, Series B*, Vol. 42, pp. 107–108.

Cervero, Robert, and Kang-Li Wu, 1998, "Subcentering and commuting: Evidence from the San Francisco Bay area, 1980–90," *Urban Studies*, Vol. 35, pp. 1059–1076.

_____, 1997, "Polycentrism, commuting, and residential location in the San Francisco Bay area," *Environment and Planning A*, Vol. 29, pp. 865–886.

Clark, Colin, 1951, "Urban population densities," *Journal of the Royal Statistical Society: Series A*, Vol. 114, pp. 490–496.

Cleveland, W. S., 1979, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, Vol. 74, pp. 829–836.

Cleveland, W. S., and S. J. Devlin, 1988, "Locally weighted regression: An approach to regression analysis by local fitting," *Journal of the American Statistical Association*, Vol. 83, pp. 596–610.

Craig, Steven G., and Pin T. Ng, 2001, "Using quantile smoothing splines to identify employment subcenters in a multicentric urban area," *Journal of Urban Economics*, Vol. 49, pp. 100–120.

Dowall, David E., and P. Alan Treffeisen, 1991, "Spatial transformation in cities of the developing world: Multinucleation and land–capital substitution in Bogotá, Colombia," *Regional Science and Urban Economics*, Vol. 21, pp. 201–224.

Fales, Raymond, and Leon N. Moses, 1972, "Land use theory and the spatial structure of the nineteenth century city," *Papers and Proceedings of the Regional Science Association*, Vol. 28, pp. 49–80.

Fan, J., 1993, "Local linear smoothers and their minimax efficiencies," *Annals of Statistics*, Vol. 21, pp. 196–216.

_____, 1992, "Design adaptive nonparametric regression," *Journal of the American Statistical Association*, Vol. 87, pp. 998–1004.

Fan, J., and I. Gijbels, 1992, "Variable bandwidth and local linear regression smoothers," *Annals of Statistics*, Vol. 20, pp. 2008–2036.

Fujita, Masahisa, and Hideaki Ogawa, 1982, "Multiple equilibria and structural transition of nonmonocentric urban configurations," *Regional Science and Urban Economics*, Vol. 12, pp. 161–196.

Gallant, A. R., 1987, "Identification and consistency in seminonparametric regression," in *Advances in Econometrics: Fifth World Congress, Volume 1*, T. F. Bewley (ed.), New York: Cambridge University Press.

_____, 1983, "Unbiased determination of production technologies," *Journal of Econometrics*, Vol. 20, pp. 285–323.

_____, 1981, "On the bias in flexible functional forms and an essentially unbiased form: The Fourier flexible form," *Journal of Econometrics*, Vol. 15, pp. 211–245.

Garreau, Joel, 1991, *Edge City*, New York: Doubleday.

Giuliano, Genevieve, and Kenneth A. Small, 1991, "Subcenters in the Los Angeles region," *Regional Science and Urban Economics*, Vol. 21, pp. 163–182.

Gordon, P., H. W. Richardson, and H. L. Wong, 1986, "The distribution of population and employment in a polycentric city: The case of Los Angeles," *Environment and Planning A*, Vol. 18, pp. 161–173.

Greene, D., 1980, "Recent trends in urban spatial structure," *Growth and Change*, Vol. 10, pp. 29–40.

Griffith, Daniel A., 1981, "Modelling urban population density in a multi-centered city," *Journal of Urban Economics*, Vol. 9, pp. 298–310.

Heikkila, E., P. Gordon, J. I. Kim, R. B. Peiser, and H. W. Richardson, 1989, "What happened to the CBD-distance gradient?: Land values in a polycentric city," *Environment and Planning A*, Vol. 21, pp. 221–232.

Ihlanfeldt, Keith R., and David L. Sjoquist, 1990, "Job accessibility and racial differences in youth employment rates," *American Economic Review*, Vol. 80, pp. 267–276.

Kain, John, 1968, "Housing segregation, Negro employment, and metropolitan decentralization," *Quarterly Journal of Economics*, Vol. 82, pp. 175–197.

Macauley, Molly, 1985, "Estimation and recent behavior of urban population and employment density gradients," *Journal of Urban Economics*, Vol. 18, pp. 251–260.

McDonald, John F., 1989, "Econometric studies of urban population density: A survey," *Journal of Urban Economics*, Vol. 26, pp. 361–385.

_____, 1987, "The identification of urban employment subcenters," *Journal of Urban Economics*, Vol. 21, pp. 242–258.

McDonald, John F., and H. Woods Bowman, 1979 "Land value functions: A reevaluation," *Journal of Urban Economics*, Vol. 6, pp. 25–41.

_____, 1976, "Some test of alternative urban population density functions," *Journal of Urban Economics*, Vol. 3, pp. 241–252.

McDonald, John F., and Daniel P. McMillen, 1990, "Employment subcenters and land values in a polycentric urban area: The case of Chicago," *Environment and Planning A*, Vol. 22, pp. 1561–1574.

McDonald, John F., and Paul J. Prather, 1994, "Suburban employment centers: The case of Chicago," *Urban Studies*, Vol. 31, pp. 201–218.

McMillen, Daniel P., 2002, "Nonparametric employment subcenter identification," *Journal of Urban Economics*, forthcoming.

_____, 1996, "One hundred fifty years of land values in Chicago: A nonparametric approach," *Journal of Urban Economics*, Vol. 40, pp. 100–124.

McMillen, Daniel P., and John F. McDonald, 1998, "Suburban subcenters and employment density in metropolitan Chicago," *Journal of Urban Economics*, Vol. 43, pp. 157–180.

_____, 1997, "A nonparametric analysis of employment density in a polycentric city," *Journal of Regional Science*, Vol. 37, pp. 591–612.

Meese, Richard, and Nancy Wallace, 1991, "Nonparametric estimation of dynamic hedonic price models and the construction of residential housing price indices," *Journal of the American Real Estate and Urban Economics Association*, Vol. 19, pp. 308–332.

Mills, Edwin S., 1972, *Studies in the Spatial Structure of the Urban Economy*, Baltimore, MD: Johns Hopkins University Press.

_____, 1970, "Urban density functions," *Urban Studies*, Vol. 7, pp. 5–20.

_____, 1969, "The value of urban land," in *The Quality of the Urban Environment*, H. Perloff (ed.), Washington, DC: Resources for the Future.

Mills, Edwin S., and Jee Peng Tan, 1980, "A comparison of urban population density functions in developed and developing countries," *Urban Studies*, Vol. 17, pp. 313–321.

Muth, Richard F., 1969, *Cities and Housing*, Chicago: University of Chicago Press.

Pavlov, A. D., 2000, "Space-varying regression coefficients: A semi-parametric approach applied to real estate markets," *Real Estate Economics*, Vol. 28, pp. 249–283.

Richardson, H. W., P. Gordon, M-J. Jun, E. Heikkila, R. Peiser, and D. Dale-Johnson, 1990, "Residential property values, the CBD, and multiple nodes: Further analysis," *Environment and Planning A*, Vol. 22, pp. 829–833.

Robinson, P. M., 1988, "Root-*N*-consistent semiparametric regression," *Econometrica*, Vol. 56, pp. 931–954.

Ruppert, D., and M. P. Wand, 1994, "Multivariate locally weighted least squares," *Annals of Statistics*, Vol. 22, pp. 1346–1370.

Schwarz, G., 1978, "Estimating the dimension of a model," *Annals of Statistics*, Vol. 6, pp. 461–464.

Shukla, Vibhooti, and Paul Waddell, 1991, "Firm location and land use in discrete urban space: A study of the spatial structure of Dallas–Fort Worth," *Regional Science and Urban Economics*, Vol. 21, pp. 225–253.

Small, K. A., and S. Song, 1994, "Population and employment densities: Structure and change," *Journal of Urban Economics*, Vol. 36, pp. 292–313.

Stone, C. J., 1977, "Consistent nonparametric regression," *Annals of Statistics*, Vol. 5, pp. 595–645.

White, Halbert, 1980, "A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica*, Vol. 48, pp. 817–838.

Yuming, Fu, and Tsur Somerville, 2001, "Site density restrictions: Measurement and empirical analysis," *Journal of Urban Economics*, Vol. 49, pp. 404–423.

Central banking and the economics of information

Edward J. Green

Introduction and summary

Advances in the economic management of information have worked pervasive change throughout the world economy and its financial system. Change due to the adoption of electronic computing and communications has been highly visible. Another kind of advance, in the design of organizations (including financial institutions) to allow them to function well in private-information environments, has been equally significant. In this article, I examine how both types of advance affect a key sector of the financial system: central banking. This examination focuses on the following three areas.

First, I study the implications of innovation induced by information technology in payment arrangements for monetary policy objectives. Some observers have suggested that such innovations, including the introduction of electronic money products or *e-money*, may nullify the relationship between issuance of money by the central bank and the price level.[1] I explain why I think that this will not happen. More precisely, I explain why innovative payment arrangements will not nullify the long-run identity between the rate of money growth and the rate of increase of the price level.

Second, I consider operational and design requirements for electronic information systems to ensure the integrity and security of the financial system (including central bank settlement of large-value payments). I conclude that an enlarged role for central banks may be warranted. Specifically, to a considerably greater extent than in the past, central bankers need to understand themselves as customers and advocates in a market for information technology, where the quality of goods and services provided by the market, rather than only the individual choices of the central bank itself, determines how high a level of integrity and security is feasible.

Third, I examine the meaning of, and requirements for, transparency of central bank decision-making. A central bank is usually said to be transparent if it makes its decisions visible to the public. For example, the Federal Reserve increased its transparency in the 1990s when it began to announce the new target for the federal funds rate immediately after the Federal Open Market Committee (FOMC) meeting in which the target was set. I suggest extending the concept of transparency beyond decisions per se to the full range of information about the central bank that is relevant to the formulation and implementation of present and future monetary policy. I argue that the decentralized structure that establishes 12 Reserve Banks as independent corporations enhances the transparency of the Federal Reserve System.

Information technology, the payments system, and monetary policy

Payments innovation as a central banking issue

Information technology is engendering rapid innovation in payment systems. E-money designed for small-value payments is the family of innovations within the past decade that is most visible to the broad public.[2] In addition, there are new technology applications that promise dramatic expansion of the applicability of netting and related procedures for large-value payments. The most conspicuous (albeit not yet operational) worldwide example is the continuous linked settlement (CLS) system, which was designed in response to market and supervisory pressure to reduce the current lag of several days between the initiation

Edward J. Green is a senior policy advisor at the Federal Reserve Bank of Chicago. *The author is indebted to Charles Goodhart and Staffan Viotti for their helpful comments on a draft of this article. This article was originally published as Edward Green, "Central Banking and the Economics of Information," in* Challenges for Central Banking, *Santomero, Viotti, Vredin (eds.), Kluwer Academic Publishers, April 2001, pp. 155–171. It has been reprinted here with the publisher's permission. The article has been edited to suit the format of this publication.*

of a foreign exchange transaction and its completion.[3] In the U.S., the Clearing House Interbank Payments System (CHIPS) has incorporated a sophisticated trade-matching algorithm in its payments system for making large-value dollar funds transfers. The new protocol makes CHIPS settlement a closer substitute for Fedwire, the Federal Reserve's real-time gross settlement system, as a way to transfer funds for most purposes.[4]

Retail e-money is more similar to these large-value payment innovations than may be apparent. As typically conceived, an electronic-money product resembles a check in being transferable prior to settlement, although in detail, the e-money is a virtual bearer security while the check is a negotiable instrument.[5] An electronic-money product resembles a credit card in giving the payee a claim on the issuer rather than on the payor or drawer. What is new about e-money, at least relative to recent payment arrangements, is the combination of these two features. While the negotiability of a check is seldom used in practice because each drawer's ability to settle is not widely established, e-money is envisioned to be routinely transferred because the issuer's ability to settle will be public knowledge. The familiar payment instrument (although no longer used in most countries) that e-money will most closely resemble, then, is a banknote.[6] Banks used to accept outside money (typically gold or silver coin, in the old days) as payment for a paper certificate, the banknote, that the bank would exchange for outside money on demand. Someone who accepted a banknote could pay it to someone else who would redeem it, and the payor and payee would thus avoid the risk and inconvenience of using outside money directly for their transaction.

One way of looking at the common use of an instrument's transferability is as a netting arrangement. The issuer or intermediary accepts a payment from the first payor and makes a payment to the last payee to use the instrument, with an arbitrary number of intervening transactions being settled without the issuer's direct involvement, but simply by transfer of the instrument. Moreover, the profitability of e-money, which derives from the issuer being able to convert money received in payment to an illiquid security with positive yield until having to make payment, relies on this netting to forestall premature liquidation of the security. From this perspective, e-money shares the same main feature of economic interest of the large-value payment innovations currently being developed.

With that preamble, I propose that increased scope for net settlement is the predominant aspect of payments system innovation that might raise a fundamental issue of monetary policy.[7] (This is not to say that it would be the only aspect relevant for other purposes, such as addressing prudential, competitive, and consumer-protection issues.) By a fundamental issue, I mean one that current understanding would not provide a good way to address. Extensive conversion of reservable deposits, or of deposits included in a monetary aggregate, to e-money might be a serious problem for monetary policy in some sense, for example, but it would not be a fundamental problem, because reformulation of the basis for computing the reserve requirement or monetary aggregate to include e-money would be an obvious solution. By this premise, a potentially open-ended examination of the implications of innovation for monetary policy can be narrowed to an examination of the monetary implications of increasing the scope of net settlement.

Several years ago, central bankers concluded that the implications of payments system innovation for the formulation and implementation of monetary policy in the context of e-money were not fundamental.[8] This conclusion was based partly on the argument that the use of money for small-value transactions is insignificant for monetary policy. That argument, of course, provides no assurance regarding innovation in large-value payments. However, there were other, independent arguments that reached the same conclusion. The central bankers suggested that, in general, as long is there is a way to induce demand for reserves, and as long as the central bank can finance open-market operations to affect the supply of reserves, conduct of monetary policy broadly along present lines should remain feasible. Inducing demand might require some regulatory distortion if interest were not to be paid on the reserves, and providing resources for open-market operations might require funding of the central bank with tax revenue (as is already done in the UK) if interest were to be paid, but even such measures were not generally viewed as serious obstacles to the conduct of monetary policy.

The question that past discussion has not resolved or even framed very explicitly is whether innovation in payment arrangements might conceivably cause the *objective* of monetary policy to change. Central bankers view themselves as making a tradeoff between the avoidance (or control) of inflation and other policy objectives such as growth and full employment. If innovation in payment arrangements were to reduce the sensitivity of the economy to inflation, then central bankers presumably should respond by giving greater weight to those other policy objectives relative to the inflation objective.

Two models of payments innovation and inflation

I discuss two economic models that purport to address the effect of payment innovation on the welfare cost of inflation. The first model suggests that innovation should reduce the welfare cost of a given level of inflation. However, the model incorporates an assumption inappropriate to studying this question. I formulate an alternative model that suggests that the welfare cost of inflation is unaffected by innovation.

I start off with a well-known model, according to which payment innovation potentially can reduce the sensitivity of the economy to inflation.[9] This model posits that transactions can be made by using either money or an alternative, nonmonetary, technology. The researchers who developed the model call the alternative technology *credit*, *financial intermediation*, or *e-money*. A buyer's decision which technology to use is based on cost minimization. The cost of using money is the interest income that is forgone by holding money rather than an interest-bearing, but illiquid, security.[10] The buyer's cost of using the alternative technology—call it *credit*—for a particular transaction depends on the seller. For any buyer, there are a few sellers—think of them as being his immediate neighbors—to whom it is very inexpensive to establish creditworthiness, a few others to whom it is very costly, and other sellers at every level in between. Each buyer desires the differentiated goods of all sellers, so he will use credit to buy from his neighbors but money to buy from distant strangers. There will be some critical distance at which buyers switch from using money to credit.

Now suppose that, for every distance, the cost of establishing creditworthiness to a seller at that distance falls by half. Then, if the seignorage tax does not change, the critical distance will double. Some payments that would have been made with money before will now be made with credit—in particular, those payments from buyers to sellers who are located farther away than the old critical distance but closer than the new one. Such a fall in the cost of establishing creditworthiness is how a payment innovation is represented in this model.

The utility loss due to a marginal increase in the price of a good is proportional to the quantity of the good consumed.[11] More generally, the utility loss due to a uniform marginal increase in the prices of a set of goods is proportional to the sum of the quantities of those goods consumed. An increase in the rate of inflation (that is, in the seignorage tax rate) translates into an after-tax price increase on all goods bought with money. The upshot is that the fewer are the goods that a buyer buys with money, the smaller is the buyer's utility loss from an increase in the rate of inflation. Since innovation in payment arrangements reduces the set of goods that each buyer buys with money, innovation reduces the aggregate welfare cost of inflation according to this analysis. An implication would seem to be that, as innovation decreases the cost of alternative payments technology and correspondingly increases its use, central bankers should care less about inflation and should turn their attention more to other policy goals.

The most satisfactory way of modeling an innovation in payment arrangements is to represent it in a general equilibrium model having the feature that people's willingness to accept money in exchange for valuable commodities arises naturally as an equilibrium phenomenon, rather than being imposed by an ad hoc constraint against goods-for-goods trades not involving money.[12] The foregoing analysis is best regarded as an essay to think in rough and ready terms about how such a fully articulated model would work. The analysis leans heavily on the premise that money is not used at all in making credit payments. In contrast, actual credit is almost always nominal, so money is essential to extinguish or settle it. Since the payment innovations that are being made on the basis of information technology are specifically means of economizing on the use of money for settlement, the purpose itself must be important for understanding the innovations. The observation that there is no debt-settlement role for money in the model just presented should be a warning bell about its appropriateness for this use.

A more appropriate proxy for a fully articulated analysis, I think, is to imagine a payment arrangement as being a protocol according to which a buyer can costlessly issue real debt (that is, make enforceable promises to provide specified quantities of goods at future dates) to finance part of a purchase, but according to which at least a specified fraction of the value of the purchase must be paid in money. This fraction corresponds to the "netting ratio," that is, to the ratio of the aggregate value of gross payments settled via the arrangement to the aggregate value of net payments made in money to effect settlement. Thus an increase in the netting ratio of the economy is a good representation of a payments innovation. Now, as a rough and ready analogy, think of the netting ratio as being just the traditional money multiplier with net payment playing the role of inside money. The clear intuition from this analogy ought to be that a payments innovation will raise the price level (by increasing the amount of inside money) but have no other effect. This intuition follows from the idea that

money is neutral in the long run, that is, that an increase in the stock of money may have transitory effects on real economic activity but will have no effect asymptotically. In particular, the long-run welfare effects of monetary policy should be identical in the post-innovation economy to what they were in the pre-innovation economy. This is the conclusion that I would expect a sound, fully articulated analysis to yield. On the basis of this expectation, I do not believe that innovation in payment arrangements constitutes a fundamental change.

Requirements for integrity and security of the financial system

Central banks have undertaken to promote the integrity and security of the financial system infrastructure, and a central bank is directly accountable for the integrity and security of its own operations.[13] Integrity means immunity from failure when operated and used even under extreme conditions (such as during a period of financial market volatility) but in good faith. Security means immunity from failure due to attempted impairment or bad-faith use by an authorized or unauthorized user.

Major components of financial system infrastructure have relied on electronic computing and communication technology for several decades. In most countries these components include, for example, the real-time gross settlement system for large-value transfers and the system of ownership registration for government securities. Other components, such as securities-market trading systems and systems for assessing and documenting the credit quality of assets intended for securitization are progressively becoming dominated by electronic technology as well.

Old-fashioned requirements of security and integrity continue to be relevant in the context of electronic information technology. For example, physical facilities have to be guarded adequately; the authorization, execution, and recording of transactions ought not to be done by the same person; and there must be sufficient investment in maintenance and redundancy of equipment to control the risk of mechanical and electrical failures.

In addition, three features of electronic technology, and of software in particular, create problems that are new or much more intense than before. First, a software defect is present in exactly the same form on all machines that run the software, so that redundancy of equipment provides no protection from such a defect. Second, there is the problem that software tends to be subject to dramatic failure on account of a defect in any one of a profusion of details. For example, in 1985, when the number of distinct issues of U.S. government securities grew too large to be represented by the address field in a program instruction (analogous to the recent century-date-change problem, in which the commonly used two-digit representation of a year ceased to be adequate), the unintended behavior of the program had business consequences that required the Bank of New York to borrow more than $20 billion at the discount window of the Federal Reserve Bank of New York.[14] More important, this episode highlighted the potential for serious disruptions to the payments system and the financial markets, although they were avoided in this instance.[15] Third, besides the problem of integrity in each of many individual components, these components—often programmed independently of one another—must interact in precisely specified ways in order to be compatible. An example of what can happen otherwise was provided last year by the Chicago Board of Trade, which temporarily had to suspend activity on its electronic trading system for financial derivatives (called Project A) because of such a system-programming problem.[16] Project A is a demonstration project being conducted by an exchange that is still mainly organized as an open-outcry trading pit. If this suspension had taken place on an exchange that relies primarily on electronic trading, as some of the world's principal exchanges already do and others envision doing soon, then there would have been an exchange-wide suspension of trading with potential implications beyond the exchange itself.

In the past, central banks and other financial intermediaries often have programmed idiosyncratic, proprietary systems suited to their individual needs. The critical need for this software to perform accurately and reliably, in view of the features that I have described above, makes such an approach increasingly risky and inordinately expensive as software becomes highly complex. The preferable approach is to synthesize a system by relying as far as possible on generic modules that are widely enough used to justify (and to share) the heavy cost of exhaustive testing, and that preferably have been used together in various combinations sufficiently often that there is a high degree of confidence in their compatibility. Besides mitigating in the first instance the problems of integrity that I have described, maintaining a system of components in widespread, current use helps to ensure that the most skilled technicians will be available (as both employees and contractors) to maintain the system and to make prompt, effective repairs when necessary.

Following this modular-design approach means depending more than previously on the general market for software and software-operated information services to meet information-technology needs. A caveat regarding this dependence has to do with the unusually high premium that financial system customers place on integrity and security. Constituting part of a niche market in this respect, the financial services industry may sometimes not be a priority customer of the software industry. The market for advanced encryption technology provides a case in point. The Digital Encryption System (DES) has been widely incorporated as a security measure in financial system software since its introduction in the 1970s. For most of that time, DES has been regarded as a commercially reasonable security measure for large-value transactions. Progressively through the 1990s, however, advances in code-breaking techniques have raised some doubt regarding the adequacy of DES encryption. A more secure encryption algorithm based on DES, the Triple Data Encryption Algorithm (TDEA, informally known as "triple DES"), has been regarded by experts for some time to offer a preferable level of protection.[17] However, despite this situation having developed somewhat predictably as code-breaking research continued and triple DES having been identified early as a reasonable response to it, the current state of the market is such that conversion of a computer system to triple-DES encryption remains a costly and managerially challenging project.[18]

A parallel situation exists with respect to software integrity. For example, current industry efforts to ensure the interoperability essential to the modular design approach envisioned above may not be stringent enough to meet fully the needs of the financial system.[19] The financial system is likely to look to central banks for leadership in working with the information technology industry and its regulators (including, perhaps, defense-related agencies charged with safeguarding communications and other economic infrastructure) to ensure that needs are met. Because the character of that industry is heavily affected by the special attributes of information technology as an economic good, bringing the needs of the financial system effectively to its attention is likely to require considerable exercise of judgment and creativity, as well as tenacity.[20]

Central bank transparency

As I mentioned at the outset, the economic management of information includes design decisions regarding the structure of institutions, as well as decisions about the employment of electronic technology for computing and communication. Central bank transparency is an issue to which both kinds of decision are relevant.

For purposes of this discussion, I call a central bank transparent to the extent that it makes public the information about itself that is relevant to the formulation and implementation of present and future monetary policy. Such information might include its objectives, its understanding (in terms of both broad concepts and specific formal models) of the structure economy, its knowledge about the current state of the economy, and its decision-making protocol.

This definition is intended to separate as clearly as possible the issue of transparency from the issue of intellectual decisiveness within the central bank itself. For example, if decision-makers within the central bank are confused or divided at a point in time regarding the significance of unexpected developments, then the public's inability to attribute a coherent view of the economy to the central bank merely reflects the true situation of the central bank, not any lack of transparency.[21]

During the past decade or two, central banks have espoused transparency to a substantial extent. One reason may be that the electorate has grown to regard this as something for which the central bank is accountable, and that central bank independence is therefore politically dependent on transparency. Another reason is that central bankers have recognized that greater transparency may favorably affect the scope of action within which they can maintain credibility while responding to macroeconomic developments. This broader scope of action may make monetary policy more effective. I am not concerned here with the justification of transparency, however, but rather with the question of how to achieve it.

Some early research on this topic (for example, Canzoneri, 1985) modeled transparency in terms of disclosure by the central bank of its information regarding the economy. The advances in information technology that have been the focus of the article thus far make such a modeling perspective less convincing than it may have previously been. A central bank does possess some private information (for example, more timely access than the public to some economic statistics compiled by the government), but large corporations—particularly multinationals—presumably possess some information that the central bank lacks. Thus, on the whole, central banks do not seem unique in point of privileged access to information or to the judgment of sophisticated market participants. Where central banks may have been unique a generation ago was in possession of techniques and equipment for

sophisticated formal modeling and forecasting, which was the province of a small community of central bank experts and university researchers. Today such econometric expertise is widely available to the public. Moreover, in some economies, the dramatic growth of financial derivatives markets and the concurrent issuance of indexed and unindexed bonds (that are approximately comparable in other respects) have generated price information that is available to the general public and have facilitated the public's direct acquisition of accurate information about expectations, especially regarding inflation.[22]

A more recent approach to analyzing transparency, taken by Faust and Svensson (1998), focuses on communication by the central bank of its objective. Faust and Svensson base their work on a modified version of a model of Cukierman and Meltzer (1986), according to which the central bank has a preferred solution to an inflation/employment tradeoff, and this preference is private information. They informally recognize that this assumption could be given more satisfactory foundations by assuming, as in Wallace (1984), that the public is heterogeneous and that monetary policy can work to the advantage of some sectors but to the disadvantage of others. The tradeoff that the central bank intends to make regarding the welfare of these sectors is its private information, which the public must infer from the subset of economic outcomes that it can observe.

My impression is that this focus on a private incentive is much closer to the gist of the actual problem of transparency than a focus on private knowledge of specific facts. Recall, however, that when I proposed a definition of transparency above, I mentioned two other domains of private information besides these two. One is the central bankers' understanding of the structure of the economy and the other is their protocol for reaching group decisions. Despite the wisdom of Faust and Svensson's decision to simplify their formal model by focusing on objectives rather than general understandings or decision protocols, transparency in these other domains is equally important and presents problems for central bankers that are at least as thorny. For example, if a central bank has a staff econometric model that is routinely discussed when monetary policy is set, should the model be described to the public and should its software even be disclosed in full? Whether or not such an initiative would do harm in any respect, I do not believe that it would give an accurate or helpful picture of the overall thinking of the monetary policy committee. The collective state of mind of such a committee would better be described by Paul Feyerabend's description of the collective state of mind of a scientific community as "a whole set of partially overlapping, factually accurate, but mutually inconsistent theories."[23] How does one accurately and informatively disclose such a state of mind to the public? I can only hope to scratch the surface of this question in this article.

There is one asymmetry in Faust and Svensson's modeling approach that would disappear if a more thoroughly *game-theoretic* approach were to be taken. That is, they present credibility as an issue of the central bank's ability to make incentive-compatible disclosure of private information, while they present transparency as an issue of the extent to which the central bank's information regarding its own objectives (or, more generally, its type in the sense that I have discussed above) is private or public. In my view, this latter information is private and the public's ability to know it is highly dependent, exactly as in the case of information about central bank actions, on there being an institutional framework that gives the central bank an incentive for accurate and informative reporting.

Discussion of institutional design to enhance transparency tends to focus on specific proposals such as the prompt publication of minutes of meetings where policy is set. My sense is that such proposals rely for their effectiveness on more fundamental structural features of the central bank. To illustrate this idea, let me cite a structural feature of the Federal Reserve that I believe plays a most significant role in achieving transparency: its decentralized structure. There are 19 persons, the seven governors of the Federal Reserve Board and the 12 presidents of the Federal Reserve Banks, who participate directly in the deliberations of the Federal Open Market Committee (FOMC), which sets monetary policy.[24] A substantial part of the ongoing analytical support of FOMC decision-making is provided by the staffs of the Board of Governors and the Federal Reserve Bank of New York. However, the fully independent participation of each Reserve Bank president is buttressed by the president's status as the head of a separately chartered corporation that comprises, among other things, a research department under the unilateral control of that president. The autonomy that is built into this structure has produced, over time, open discussion of a number of policy foundations and alternatives that I believe might have received less or later exposure in a more centralized institutional framework.

Several important examples from recent decades support this case. Beginning in the 1960s, the Federal Reserve Bank of St. Louis conducted a sustained program of research and advocacy regarding the control

of monetary aggregates as a basis for conducting monetary policy.[25] In the 1970s and 1980s, the Federal Reserve Bank of Minneapolis played a significant role in developing *general equilibrium* monetary models for policy analysis as an alternative to the macroeconomic modeling approach that was then dominant in the Federal Reserve.[26] In the early 1990s, the Federal Reserve Bank of Cleveland persistently made a case that the benefits of bringing inflation under control would not be fully garnered until exact price stability had been achieved.[27] These essays in analysis and persuasion have been both more vigorous and more open to public scrutiny than I believe they would have been if they had been led by policymakers of equivalent seniority, but operating within a more hierarchically organized central bank. In all three cases, the advocates of heterodox positions within the central bank have had to depend heavily on informed public opinion, and particularly on the endorsement of economists in the academic community, to affirm the correctness of their views. Thus, the decentralized design of the U.S. central banking system systematically forces policy debate out into the open marketplace of ideas, to the benefit of both the transparency of the Federal Reserve System and the intellectual caliber of the discussion. The history of the three initiatives that I have mentioned, and of others as well, suggests that this process succeeds in identifying and evaluating significant new ideas and, where merited, progressively infusing them into the policymaking of the central bank as a whole, albeit usually not in the uncompromising form that they initially tend to be proposed. In my view, this sort of institutional design for the central bank is an important complement to the various, specific regulations (regarding, for example, the exact timing and format of public release of minutes of policy-setting meetings) that are usually recommended as means to achieve transparency and to ensure that monetary policy is publicly accountable.

Other design approaches, adopted by various central banks in recent years, have analogous roles in providing transparency. The common feature of these approaches is that, rather than attempting to achieve transparency by mandate, they set in place systems of incentives that result in an institutional culture of transparency.[28] Both a conducive culture and a clear public mandate have a place in achieving transparency. Indeed, for the central bank to have an appropriate institutional culture is probably a necessary condition for a mandate to be effective.

Conclusion

Recent, dramatic innovations in the economic management of information, and particularly their application in the payments system, might seem potentially to change the nature of central banking. On close examination, however, these developments do not significantly change the role or responsibilities of a central bank. They do not render obsolete the established body of knowledge regarding what constitutes a well designed central bank and sound central banking practice.

Similarly, intellectual advances in understanding how organizations should optimally be designed reinforce established thinking about how a central bank should be designed to achieve transparency. Indeed, these advances provide a clearer understanding of how the decentralized structure of the Federal Reserve System contributes to the effectiveness of the U.S. central bank and to the public welfare.

The one area where innovations in information technology do seem to call for new understanding is in the involvement of central banks with the technology itself. Such involvement is required to discharge both oversight and operational responsibilities. On behalf of the financial system, as well as on its own behalf, a central bank must manage problems that are rooted in the structure of the information-technology industry. Adept management is required to maintain the integrity and security of a financial system that, because of its scope and complexity, is critically dependent on information technology for its functioning.

NOTES

[1]Money issued by the central bank is known as *outside money.* Commercial banks and other such depository institutions also issue money, in effect, when they make loans. This is known as *inside money.* A requirement that depository institutions must hold reserves of outside money constrains their ability to create inside money. Reserve requirements in the U.S. and some other countries are *deposit reserves* based on the value of deposits that a depository institution holds, and in other countries are *clearing balances* based on the value of payments that a depository institution makes on behalf of its depositors. In this article, "money" means outside money unless otherwise indicated, and "reserves" is used as a generic term for either deposit reserves or clearing balances.

[2]E-money refers to a family of payment methods that include stored-value cards and "Internet cash" designed for widespread use. Payment methods designed for convenient purchasing from a single seller, such as the fare cards issued by some public transit authorities, are not within the meaning that is usually intended.

[3]The CLS system is described in Bank for International Settlements (1998).

[4]De Santis (1998) describes the proposed system and its risk-management implications. Marjanovic (1998) also provides a brief description.

[5]A negotiable instrument is one that has a particular, named individual as its beneficiary, but that allows that beneficiary to designate another person as beneficiary instead (typically, as payment for a good or service received from the new beneficiary). A bearer security is a financial instrument, such as currency, whose beneficiary is whoever happens to possess it. Some types of e-money differ from a literal bearer security in point of requiring proof of ownership beyond physical possession. That difference is not material to the analogy drawn here.

[6]That is, a bearer security issued by a bank and redeemable for coin or other legal money. Wallace (1986) and Summers and Gilbert (1996) have previously emphasized this analogy.

[7]Friedman (1999) and King (1999) suggest that extensive use of information technology might make it possible in principle for the private sector to operate a comprehensive settlement network that would be wholly outside the influence of the central bank. In that case, my premise would be violated. The gist of my argument in this section is that, although it might seem that being able to settle large gross payments with much smaller net payments is tantamount to the situation that Friedman and King have in mind, the implications for monetary policy may be materially different.

[8]Bank for International Settlements (1996) reflects some of that discussion.

[9]The model is an elaboration of a cash-in-advance model of Lucas and Stokey (1987) in which, for each trader, some goods are *credit goods* that are exempt from the cash-in-advance constraint that holds for the remaining *cash goods*. The elaboration is to endogenize the cash/credit distinction as explained here. Such models were introduced by Schreft (1992) and Aiyagari, Braun, and Eckstein (1998).

[10]This foregone interest is *seignorage* that is captured by the government, which would have to issue interest-bearing debt to finance expenditure if people would not accept money. Seignorage is thus an implicit tax on holding money.

[11]This is Roy's identity; see Deaton and Muellbauer (1980), p. 40. This principle is highly intuitive. For example, if someone purchases five daily newspapers and one weekly newsmagazine a week, then a penny increase in the price of a newspaper hurts five times as much as a penny increase in the price of a magazine. This is evidently true if consumption does not change. If the reader cuts down to four newspapers a week, then he was getting just one penny's worth of utility from the fifth paper beyond what alternative expenditure of its price would have yielded (since he elected to give it up when an extra penny was charged), so (on the simplifying assumption that utility is measured in whole penny's-worth units) he still loses a penny's worth of utility despite changing his budget allocation.

[12]Freeman (1996a, b, and 1999) and Green (1997) analyze central bank operations, and also clearinghouse operations closely akin to netting, in this way.

[13]Such a responsibility is widely conferred to, and considered an appropriate role for, central banks. Consensus to this effect is reflected, for example, in a series of documents issued under the auspices of the Bank for International Settlements during the past decade. Payment arrangements operated by private financial intermediaries (either directly or via jointly owned subsidiaries) evidently cannot be supervised in isolation from the sponsoring intermediaries themselves. Therefore, oversight of such arrangements involves coordination between the central bank and the supervision authorities for various types of intermediaries.

[14]Statements of Paul A. Volcker, Chairman, Board of Governors of the Federal Reserve System, and J. Carter Bacot, Chairman and Chief Executive Officer of the Bank of New York and the Bank of New York Company, Inc., before the Subcommittee on Domestic Monetary Policy of the Committee on Banking, Finance, and Urban Affairs, House of Representatives, December 12, 1985. Reprinted in U.S. Congress (1986).

[15]Statement of Paul A. Volcker, Chairman, Board of Governors of the Federal Reserve System, before the Subcommittee on Domestic Monetary Policy of the Committee on Banking, Finance, and Urban Affairs, House of Representatives, December 12, 1985. Reprinted in U.S. Congress (1986).

[16]Chicago Board of Trade (1999).

[17]In 1998, the Accredited Standards Committee X9 on Financial Services, a U.S. financial services industry committee working under the aegis of the American National Standards Institute (ANSI), adopted standard ANSI X9.52, which specifies triple DES as an interim encryption method for large-value financial transactions until a more durable method can be developed. In 1999, the X9 Committee issued technical guideline TG-25-1999, which expresses a consensus that (single) DES encryption no longer provides adequate security for large-value transactions. An analogous guideline for U.S. government applications and public recommendation for nongovernmental applications, "Data encryption standard (DES)," Federal Information Processing Standards (FIPS) Publication 46-3, was issued by the U.S. Department of Commerce/National Institute of Standards and Technology (NIST) in 1999. This guideline designates triple DES as an encryption algorithm of choice and permits DES for legacy systems only. The NIST is also in the process of adopting an Advanced Encryption Standard that will co-exist with, and eventually supplant, triple DES.

[18]To provide adequate electronic security, conversion to triple DES must be done in conjunction with other measures such as setting up a cryptographic key-management system. For non-U.S. entities, a further complication recognized by NIST FIPS-46-3 is that export of encryption systems deemed to provide commercially reasonable security for large-value financial applications is subject to U.S. export-control regulation.

[19]Summers (1999) raises this issue.

[20]Arrow (1974) provides a classic introduction to the issues of market structure engendered by characteristics of information-related industries.

[21]Transparency is relevant in other contexts besides monetary policy, but I have qualified the definition in this respect to screen off issues that involve special considerations (such as whether "constructive ambiguity" about the extent of the central bank's willingness to provide emergency credit is a justifiable strategy to deter implicitly subsidized risk-taking).

[22]Roughly speaking, the interest rate premium of an unindexed bond above an indexed bond of the same maturity provides a measure of expected inflation from the present to the maturity date. Techniques to extract information regarding expectations from asset-price data are surveyed by Soderlind and Svensson (1997).

[23]Feyerabend (1978), p. 39.

[24]All but one of the Reserve Bank presidents are voting members only in rotation, but their participation in deliberations is continuous.

[25]See Andersen and Carlson (1974).

[26]See Miller (1994), which is a collection of papers reprinted from the Federal Reserve Bank of Minneapolis, *Quarterly Review.*

[27]See Gavin (1991).

[28]Having such incentive systems in place is a very helpful means to attract capable, principled, intellectually independent persons to serve on the governing board of the central bank. It is also helpful to guarantee the ability of the governing board to hire and retain staff with those characteristics. Perception of the high character and ability of the central bank leadership and staff seems to contribute materially to public support for central bank independence.

REFERENCES

Aiyagari, Rao, R. Anton Braun, and Zvi Eckstein, 1998, "Transaction services, inflation, and welfare," *Journal of Political Economy*, Vol. 106, pp. 1274–1301.

American National Standards Institute, Accredited Standards Committee X9 on Financial Services, 1999, "A guide to understanding data remanence in automated information systems," technical guideline, No. TG-25-1999, Washington, DC.

_____, 1998, "Triple data encryption algorithm modes of operation," ANSI standard, No. X9.52, Washington, DC.

Andersen, Leonall, and Keith Carlson, 1974, "The St. Louis model revisited," *International Economic Review*, Vol. 15, pp. 305–327.

Arrow, Kenneth J., 1974, *The Limits of Organization*, New York: W. W. Norton.

Bank for International Settlements, 1998, "Reducing foreign exchange settlement risk: A progress report," Basel, Switzerland.

_____, 1996, "Implications for central banks of the development of electronic money," Basel, Switzerland.

Canzoneri, Matthew B., 1985, "Monetary policy games and the role of private information," *American Economic Review*, Vol. 75, pp. 1056–1070.

Chicago Board of Trade, 1999, press release, August 17.

Cukierman, Alex, and Allan H. Meltzer, 1986, "A theory of ambiguity, credibility, and inflation under discretion and asymmetric information," *Econometrica*, Vol. 54, pp. 1099–1128.

Deaton, Angus, and John Muellbauer, 1980, *Economics and Consumer Behaviour*, Cambridge, UK: Cambridge University Press.

De Santis, Vincent, 1998, "CHIPS update," presentation to the Money Transfer 98 conference sponsored by Bank Administration Institute.

Faust, Jon, and Lars E. O. Svensson, 1998, "Transparency and credibility: Monetary policy with unobservable goals," Board of Governors of the Federal Reserve System, international finance discussion paper, No. 605.

Feyerabend, Paul, 1978, *Against Method*, London: Verso Editions.

Freeman, Scott, 1999, "Rediscounting under aggregate risk," *Journal of Monetary Economics*, Vol. 43, pp. 197–216.

_____, 1996a, "Clearinghouse banks and banknote over-issue," *Journal of Monetary Economics*, Vol. 38, pp. 101–115.

_____, 1996b, "The payments system, liquidity, and rediscounting," *American Economic Review*, Vol. 86, pp. 1126—1138.

Friedman, Benjamin, 1999, "The future of monetary policy: The central bank as an army with only a signal corps?," National Bureau of Economic Research, working paper, No. 7420.

Gavin, William T., (special issue ed.), 1991, "Price stability: A conference sponsored by the Federal Reserve Bank of Cleveland, November 8–10, 1990," *Journal of Money, Credit, and Banking*, Vol. 23, pp. 433–631.

Green, Edward J., 1997, "Money and debt in the structure of payments," *Bank of Japan Monetary and Economic Studies,* Vol. 15, pp. 63–87, reprinted in 1999, *Quarterly Review*, Federal Reserve Bank of Minneapolis, Vol. 23, pp. 13–29.

King, Mervyn, 1999, "Challenges for monetary policy, new and old," presentation to the New Challenges for Monetary Policy symposium of the Federal Reserve Bank of Kansas City, Jackson Hole, WY, August 26–28.

Lucas, Robert E., Jr., and Nancy L. Stokey, 1987, "Money and interest in a cash-in-advance economy," *Econometrica*, May, Vol. 55, No. 3, pp. 491–513.

Marjanovic, Steven, 1998, "N.Y. Clearing House in final leg of overhaul," *American Banker,* Vol. 163, No. 72, May 17, p. 1.

Miller, Preston J., (ed.), 1994, *The Rational Expectations Revolution: Readings from the Front Line*, Cambridge, MA: Massachusetts Institute of Technology Press.

National Institute of Standards and Technology, 1999, "Data encryption standard (DES)," U.S. Department of Commerce, report, No. FIPS-46-3.

Schreft, S. L., 1992, "Transaction costs and the use of cash and credit," *Economic Theory*, Vol. 2, pp. 283–296.

Soderlind, Paul, and Lars Svensson, 1997, "New techniques to extract market expectations from financial instruments," *Journal of Monetary Economics*, Vol. 40, pp. 383–429.

Summers, Bruce, 1999, "Integrity and trust in electronic banking," presentation to the 1999 Software Engineering Symposium of the Carnegie Mellon Software Engineering Institute.

Summers, Bruce, and R. Alton Gilbert, 1996, "Clearing and settlement of U.S. dollar payments: Back to the future?," *Review*, Federal Reserve Bank of St. Louis, Vol. 78, pp. 3–27.

U.S. Congress, 1986, "The Federal Reserve Bank of New York discount window advance of $22.6 billion extended to the Bank of New York," serial, No. 99-65.

Wallace, Neil, 1986, "The impact of new payment technologies: A macro view," in *Technological innovation, regulation, and the monetary economy*, C. Lawrence and R. P. Shay (ed.), Cambridge, MA: Harper & Row, pp. 201–206.

_____, 1984, "Some of the choices for monetary policy," *Quarterly Review*, Federal Reserve Bank of Minneapolis, Vol. 8, pp. 15–24, reprinted in Miller, 1994.

Competition among banks: Good or bad?

Nicola Cetorelli

Introduction and summary

In recent years we have witnessed a substantial convergence of research interest and the opening of a debate on the economic role of market competition in the banking industry. The need for such a debate may seem unjustified at first. The common wisdom would hold that restraining competitive forces should unequivocally produce welfare losses. Banks with monopoly power would exercise their ability to extract rents by charging higher loan interest rates to businesses and by paying a lower rate of return to depositors. Higher lending rates would distort entrepreneurial incentives toward the undertaking of excessively risky projects, thus weakening the stability of credit markets and increasing the likelihood of systemic failure. Higher lending rates would also limit firms' investment in research and development, thus slowing down the pace of technological innovation and productivity growth. Lower supply of loanable funds, associated with higher lending rates, should also be reflected in a slower process of capital accumulation and, therefore, in a lack of convergence to the highest levels of income per capita.

These are some of the conventional effects that market power in the banking industry is commonly thought to generate. However, in more recent years, researchers have begun analyzing additional issues in the matter of bank competition, highlighting potentially negative aspects and so raising doubts regarding the overall beneficial welfare impact of bank competition on the economy. The research effort devoted to this issue has picked up noticeably, a sign that the time is ripe for an open debate regarding the costs and benefits of bank competition.[1]

The policy implications associated with this issue, related to the regulation of the market structure of the banking industry, are especially relevant. In fact, banking market structure is a traditional policy variable for the regulator. Implicitly or explicitly motivated by the desire to restrain banks' ability to extract rents, policymakers would typically recommend measures aimed at fueling competition, promoting the liberalization of financial markets and removing barriers to entry (see, for example, Vittas, 1992). In light of the most recent regulatory changes affecting the U.S. financial industry, the policy relevance for U.S. regulators is more current than ever. In 1992 intrastate branching restrictions were relaxed, followed in 1994 by the passage of the Riegle–Neal Interstate Banking and Branching Efficiency Act, which allows bank holding companies to acquire banks in any state and, as of June 1, 1997, to branch across state lines. Finally, 1999 saw the passage of the Financial Services Modernization (Gramm–Leach–Bliley) Act, allowing the operation of commercial banking, investment banking, and insurance underwriting within the same holding company. Such regulatory changes continue to have a significant impact on the market structure of the banking industry and on banks' competitive conduct. A deeper analysis of the economic role of bank competition should thus contribute to our understanding of the role of the regulator and the consequences of regulatory action and, therefore, support more effective policymaking.

The goal of this article is to summarize some of the arguments that have recently emerged and to suggest some new lines of investigation. In the next section, I describe theoretical contributions that have identified both positive and negative effects of bank competition. Subsequently, I illustrate the results of

Nicola Cetorelli is an economist at the Federal Reserve Bank of Chicago. *The author thanks David Marshall for his useful comments.*

existing empirical studies, which present mixed evidence regarding the economic role of bank competition. The main conclusion that seems to emerge from the review of the current literature is that the market structure of the banking industry and the related conduct of banking firms affect the economy in a much more complicated way than through the simple association: more market power equals higher lending rates and lower credit quantities. By combining the various research studies, I identify multiple effects of bank competition, acting along different economic dimensions, suggesting the existence of tradeoffs and leading us toward more sophisticated normative considerations associated with bank competition. For example, as I describe in detail later, there is evidence from recent work to support the conventional wisdom that a more concentrated banking industry imposes a deadweight loss in the credit market as a whole, resulting in a reduction in the total quantity of loanable funds and slower economic growth. However, the effect appears to be *heterogeneous* across industry sectors, and younger firms in industries that are heavily dependent on banks for investment funds actually seem to grow faster if they deal with a concentrated banking sector.

The final section of the article explores some additional lines of research on the economic role of bank competition. For instance, does it matter whether banks are government owned? To what extent do government-owned banks behave differently from privately owned banks? Could common ownership across different government-owned banks imply a cartel-like behavior?

A separate question is whether the role of bank competition varies depending on how restrictive is the regulatory environment of the banking sector. Banks may be or may not be allowed, for example, to own and control nonfinancial firms, or to engage in securities or insurance underwriting and selling, or real estate investments. The possibility for banks to be active in multiple markets and face competition from nonbank firms in such markets may have an impact on the role of bank competition in the economy. For example, the possibility to offer a wider array of products and services may allow banks to "capture" and retain clients even while facing intense competition in traditional banking markets.

Finally, another dimension of analysis is the exploration of the possible relationship between market power in the banking industry and that in other sectors of the economy. Does a concentration of market power in banking lead banks to extend credit to few firms, which grow in size and make their sectors concentrated, or rather does bank concentration promote the continuous entry of firms, thus contributing to lower concentration in other industries? Theoretical conjectures could be presented suggesting either effect.

I present an illustration of these separate lines of inquiry and some empirical evidence. This evidence confirms that the market structure of the banking industry and the related conduct of banking firms have an important role in maintaining a well-functioning economy and that normative action regarding bank regulation requires careful consideration.

Theoretical arguments

I begin by illustrating the most common theoretical arguments used to identify positive and negative economic effects of bank competition. In a stylized model of economic growth, Pagano (1993) showed that market power, by allowing banks to charge higher loan rates and compensate savers with lower deposit rates, does indeed reduce the equilibrium quantities of funds available for credit, hence generating a direct negative effect on the rate at which the economy can grow. Guzman (2000) confirms this negative effect of market power in a general equilibrium model of capital accumulation. He compares two identical economies, one with a monopolistic bank and the other with a competitive banking sector, and shows that the monopolistic bank produces an unequivocally depressing effect on capital accumulation for two possible reasons. If the conditions exist for credit rationing, quantities are rationed more with a monopolistic bank than within a competitive setting. Without credit rationing, monopoly power in banking is still inefficient because it leads to excessive monitoring. As Guzman argues, this is due to the fact that with monopoly power loan rates are higher, and with higher rates the likelihood of default also increases (moral hazard). Consequently, the monopolistic bank has to sustain a higher cost to monitor entrepreneurs, thus diverting resources that could otherwise be available for lending.

In perhaps the most widely cited article about this issue, Petersen and Rajan (1995) focus on the role banks play in financing new businesses. In a stylized theoretical model, the authors show that young firms with no record of past performance may actually receive more credit, and at better rates, if they are in a market where banks have monopoly power. The intuition is the following. Lenders facing a pool of risky (because yet unknown) borrowers should incorporate an appropriate premium in their lending rates to cover a likelihood of default potentially higher than that among already established entrepreneurs. Consequently, lending rates for this category of borrowers should be high and credit partially rationed. However, in such a

scenario, a bank with market power has an alternative lending strategy. It can charge "introductory" lower rates, attract more—and possibly on average better—young entrepreneurs, and establish a *lending relationship* with them, with the prospect of extracting rent (charging higher rates) in the future from those who are eventually successful. This strategy of initial "subsidization" and subsequent "participation" in successful firms' profits is feasible if the bank has market power. The bank relies on the fact that the successful firms will not be bid away by competitor banks in the future. On the other hand, in a competitive market, a bank sustaining the initial cost of offering credit at a lower rate could not count on its ability to retain the successful customers.

In a more recent paper, Shaffer (1998) points out another possible shortcoming associated with bank competition. One of banks' main functions is that of performing screening, separating prospective entrepreneurs by quality categories. Shaffer shows that the average quality of a bank's pool of borrowers declines as the number of competitors in the market increases. The intuition is based on the possibility that banks' screening technologies may not accurately report the borrowers' true characteristics. Suppose the screening model used by banks is indeed imperfect, in the sense that with a certain probability entrepreneurs of high quality can be identified as being of low quality, and vice versa. Also assume that a bank cannot distinguish between a new loan applicant and someone who has already been denied credit by another institution. As a result, rejected applicants (either of high or low quality) can continue to apply to other banks; the more banks there are in the market, the higher the likelihood that a low-quality applicant receives credit. This occurrence is known as "winner's curse": A bank that agrees to extend a loan may be "winning" the right to fund a lemon.

Also focusing on bank screening, Cao and Shi (2000) argue that, because an increase in the number of banks operating in the market exacerbates the winner's curse, the number of banks active in performing screening and competing in supplying credit would actually fall; as a result, loan rates would be higher and credit quantities smaller than in a market with fewer banks.

Dell'Ariccia (2000) explores another model of bank screening, showing that as the number of banks increases, the likelihood that banks will actually screen entrepreneurs, as opposed to lending indiscriminately, decreases. His argument is based on the observation that entrepreneurs may be averse to being screened. For instance, the screening process may be time-consuming and in the process the firm may miss profit opportunities. Alternatively, an entrepreneur may not want to reveal the true creditworthiness of the project. In slow-growth periods or during recessions, screening may be the optimal strategy, since there is a high probability that entrepreneurs demanding credit may be of low quality and have already been rejected by other banks. However, in periods of economic expansion, when there is a higher proportion of new, untested entrepreneurs, banks competing for market share may choose to offer lending contracts involving no screening. The interesting result is that by bearing a higher risk in the upswing of the economic cycle, banks are more likely to plant the seeds for a subsequent recession.

Manove, Padilla, and Pagano (2000) observe that screening and collateralizing are substitutes from the point of view of a bank's lending strategy. A bank screens to select high-quality entrepreneurs and reduce the risk of default among low-quality ones. However, if an entrepreneur posts full collateral, then the bank may not have an incentive to screen (where screening is a costly activity) since the bank would be protected in the event of default. Consider a world with high- and low-quality entrepreneurs, where high-quality ones have a higher probability of picking a good project. Entrepreneurs know whether they are of high or low quality. In a competitive banking market, banks would offer loans only to those entrepreneurs whose projects were screened and thus recognized as successful. However, banks have to offer them credit at a rate high enough to recoup the total cost of screening (including the screening cost component of the entrepreneurs whose applications were rejected). High-quality entrepreneurs can separate themselves from the pool by offering to post collateral on their loan (low-quality entrepreneurs would not offer to post collateral, since they face a high probability of loss if their project turns out to be unsuccessful). Hence, banks would only screen low-quality borrowers and extend credit to those who were able to pick a successful project. All high-quality entrepreneurs (some of which will still be unsuccessful) would receive credit since they posted collateral and, therefore, do not constitute a risk for the bank.

What happens in a market with a monopolistic bank? According to the authors, such a bank may not have the incentive to accept collateral from high-quality entrepreneurs. This is because the monopolistic bank is able to appropriate the surplus generated by successful projects. Hence, for this bank, screening implies a higher rate of return and, therefore, may be preferred to accepting collateral. In this case, the

monopolistic bank screens all projects, thus eliminating the allocation of resources to entrepreneurs destined to fail.

Multiple effects of banking market structure

If banks' role were simply that of intermediating between supply and demand of credit, then market power in the hands of banks could only generate the conventional negative effect associated with rent extraction. However, banks fulfill other important functions—in particular their role in screening prospective entrepreneurs and in allocating capital resources to the best social uses. The studies described above share the insight that market competition may *distort* banks' incentive to perform these additional roles. A legitimate observation, therefore, is that banking market structure produces multiple effects, (and of opposite directions) on the economy. On the one hand, market power may enable banks to extract rents and distort the equilibrium of the credit market away from one where the *quantity* of funds supplied is the highest. On the other hand, market power may be necessary to allow banks to achieve an efficient allocation of funds, thus enhancing the *quality* of the pool of selected entrepreneurs.

The identification of a tradeoff between quantity of credit made available in the market and banks' role in allocating funds efficiently is an important insight that emerges from the most recent analysis of bank competition. Cetorelli (1997) and Cetorelli and Peretto (2000) identify both roles and model the tradeoff. Both papers analyze the role of banking market structure for an economy's path of capital accumulation and growth using a dynamic, general equilibrium framework. The first paper compares only two benchmark economies, one with perfect competition and the other with a monopolistic bank, while the second analyzes banks in a fully specified *Cournot* oligopoly model. The Cournot model has the nice feature that competition and monopoly are the two extremes of a continuum of market structures, wherein market power is fully captured by the number of firms if the model is symmetric, or corresponding measures of market concentration if the model is asymmetric. Cetorelli and Peretto (2000) analyze N banks competing with each other in gathering individual savings and in loaning funds to entrepreneurs. Banks have access to a screening technology that, at a cost, allows them to discriminate between high- and low-quality entrepreneurs. While the outcome of the screening test may not be observable by third parties, competitor banks can extract information about the screened entrepreneurs by simply observing whether the bank extends or denies the loan.[2] In other words, there is an informational externality that generates a free-riding problem, which weakens banks' incentives to incur the cost of screening and to carry out an information-based (efficient) lending strategy. Cetorelli and Peretto's model shows that the bank's optimal strategy entails screening entrepreneurs only with some probability, and thereby extending both "safe" (screened) and "risky" (unscreened) loans. The credit market is thus endogenously segmented: A fraction of entrepreneurs are always screened, with credit extended only to those of high quality, while the remaining proportion of entrepreneurs receive credit indiscriminately, regardless of characteristics of quality. The relative size of these two components of the credit market evolves along, and has feedback into, the path of economic development. Within this theoretical framework, two major effects of banking market structure on economic growth are identified. On the one hand, the fewer the number of banks, the smaller the total quantity of credit available to entrepreneurs, exactly as conventional wisdom suggests. On the other hand, the fewer the number of banks, the greater the incentive for banks to screen and, consequently, the larger the proportion of funds that is allocated efficiently to high-quality entrepreneurs.[3] Therefore, the number of banks governs the tradeoff between the overall size of the credit market and its efficiency. The size and efficiency of the credit market, in turn, determine the return to capital accumulation and, therefore, to saving. The main result of this model is that, because of this tradeoff, the relationship between banking market structure and steady-state income per capita may not be monotonic. In other words, the market structure that maximizes economic development is neither a monopoly nor perfect competition, but an oligopoly.

Empirical evidence

Simultaneously with the development of the theoretical debate, researchers have also begun to investigate empirically the economic role of banking market structure. As with the theoretical contributions, the empirical findings suggest that banking market structure has both positive and negative economic effects, and it is hard to establish which one ultimately dominates. For example, a few studies provide evidence of a clearly negative role of banking market power. Shaffer (1998) uses data on household income growth between 1979 and 1989 in U.S. metropolitan statistical areas (MSAs). He finds that, after controlling for other determinants of income growth, household income grows faster in MSAs with a higher number

of banks. Black and Strahan (2000) focus instead on the impact of banking market structure in fostering entrepreneurial activity. Looking at cross-industry, cross-state U.S. data, they find that the number of new firms and the number of new business incorporations are smaller in states where bank concentration is higher. Jayaratne and Strahan (1996) estimate the effect of the removal of U.S. bank branching restrictions on state income growth. The removal of such barriers should presumably enhance competition. They find that both personal income and output growth accelerated after states implemented the regulatory change. Hence, their findings suggest, indirectly, a positive effect of bank competition on economic growth.

At the same time, however, some empirical contributions have suggested a positive effect of bank concentration. For example, Petersen and Rajan (1995) analyze credit availability for a cross-section of U.S. small businesses located in markets characterized by different degrees of banking concentration. They find that firms are less credit constrained if they are in more concentrated markets. In addition, they find that younger firms pay lower loan rates in markets with higher bank concentration. Shaffer (1998), cited earlier, also finds evidence of higher loan chargeoff rates in MSAs with a higher number of banks. Collender and Shaffer (2000) report evidence that while the effect of bank concentration on household income in U.S. metropolitan areas was negative between 1973 and 1984, it was positive during the 1984–96 period. Bonaccorsi and Dell'Ariccia (2000) analyze cross-industry, cross-province Italian data and find that the rate of creation of new firms is higher in provinces with a more concentrated banking sector (an Italian province is roughly equivalent to a U.S. metropolitan statistical area). In fact, the effect is especially strong on new firms belonging to industry sectors that can be considered more informationally opaque, that is, where the technologies adopted are such that banks need to put more effort into screening and selecting entrepreneurs.

Evidence on multiple effects of banking market structure

Empirical evidence of both a positive and a negative channel through which banking market structure may affect an economy, implied by the various theoretical contributions and hinted at by the empirical evidence surveyed in the previous section, is confirmed in Cetorelli and Gambera (2001). They test the role of banking market structure using data on the growth of 36 industry sectors in 41 different countries, both developing and developed economies, expanding on the existing and well-established methodologies employed in the literature on finance and growth. The main stylized fact recognized in this literature is that a well-developed banking sector has an important, causal role in economic growth. The basic question asked in Cetorelli and Gambera is then, for a given level of development of the banking sector, what is the role of its market structure? They begin by evaluating whether countries with higher bank concentration are characterized by higher or lower growth across industry sectors. Given the opposing theoretical views described earlier, the answer to this question is not obvious. On the one hand, if bank concentration simply results in lower credit availability, then growth across industries should be slower in countries with a more concentrated banking market. On the other hand, if the market power associated with bank concentration generates positive effects, according, for example, to the relationship-lending argument of Petersen and Rajan (1995), then growth should be faster in countries with a concentrated banking sector. Cetorelli and Gambera find that bank concentration has a *negative* effect, on average, on industry growth.

However, Cetorelli and Gambera's empirical study goes beyond the analysis of this average effect of banking market structure. Using industry-specific information about the intensity with which industry sectors are dependent on external sources of finance, they perform more refined empirical tests. Rajan and Zingales (1998) constructed such an industry-specific measure of dependence, arguing that, due to idiosyncratic factors, different industry sectors are more or less in need of external sources of funding to finance capital investment. Sectors adopt different technologies, which imply different initial project scales, different gestation periods and cash harvest periods, and different reinvestment requirements. Intuitively, sectors like tobacco or leather generate large amount of funds internally that can be used for investment purposes. At the opposite extreme, sectors like computers or pharmaceuticals, characterized by uncertainty in the timing and in the rate of return of their investments, will be much more dependent on external sources of funds. Moreover, within a sector, the intensity of external financial dependence will also differ across firms of different age, with younger firms more in need than mature ones. Cetorelli and Gambera use this information to test whether banking market structure has a heterogeneous effect across industry sectors. Given the opposing theoretical views, one might argue, on the one hand, that firms in sectors especially dependent on external finance would suffer more, and therefore grow less than average, in a country with a concentrated banking sector. On the other hand, if bank concentration has positive effects, then firms in industries especially dependent on external finance

should benefit disproportionately more when faced with a concentrated banking sector. Cetorelli and Gambera's results show that industries more dependent on external finance in fact grow relatively *faster* in those countries where the banking sector is *more* concentrated. The effect is more pronounced for younger firms than for mature firms.

Cetorelli and Gambera's two main findings taken together thus confirm the existence of multiple effects of banking market structure. A more concentrated banking industry does impose a deadweight loss in the credit market as a whole, resulting in a reduction in the total quantity of loanable funds, exactly as conventional wisdom would suggest. However, the effect appears to be *heterogeneous* across industry sectors, and younger firms in industries that are heavily dependent on banks for investment funds seem to benefit from a concentrated banking sector.

New dimensions of analysis

These findings about the economic role of bank competition draw a picture regarding the normative implications for regulatory action that is much less clear than what has been suggested by conventional wisdom. In particular, it is not clear whether competition is necessarily welfare improving. Perhaps the major insight we have gained is that policy action related to bank competition needs to be coordinated across multiple dimensions. There may be more funds available in a competitive credit market, but there may also be higher rates of default and, consequently, greater waste of resources.[4] Some of these dimensions of the analysis are dependent on each other. For example, from the last section we learned that depending on the level of concentration of the banking industry, ceteris paribus, individual sectors will grow at different speeds. Therefore, banking market structure plays an important role in shaping the cross-industry size distribution within a country. Consequently, we have identified an interesting connection between regulation of the financial industry and industry planning. In addition, we find that bank concentration plays a more substantial role in growth by facilitating credit access of younger firms. To the extent that investment of younger firms is more likely to introduce innovative technologies, regulators face an unexpected tradeoff between the generally desirable effects of bank competition and the promotion of technological progress.

Government ownership

Next, I explore some additional lines of research on the economic role of bank competition. For instance, does it matter whether banks are government owned? How does government ownership affect the relationship between bank competition and industry growth discussed above? La Porta et al. (2000) have recently shown that government ownership in banking is a pervasive phenomenon observed across countries, more so in developing economies. The presumption is that public banks are less efficient and perform a poorer job in allocating capital to the best uses. The authors confirm this presumption by showing that countries where government-owned banks are more predominant are also characterized by lower rates of growth in per capita income and in productivity. In addition, these countries face slower development in financial markets. Cetorelli and Gambera (2001) test whether the degree of government ownership in banking affects the role of bank concentration in industry growth. They show that in countries with both high bank concentration and a high degree of government ownership of banks (as a proportion of total bank assets), the positive role of bank concentration on the growth of sectors highly dependent on external finance vanishes. What is left is evidence of the standard inefficiencies associated with market concentration. The positive role of bank concentration described earlier supports the argument that market power is needed for banks to be willing to efficiently screen entrepreneurs and establish lending relationships with them. The fact that no positive role for bank concentration is found in countries with high bank government ownership is consistent with the argument (see La Porta et al., 2000) that government banks are more likely to be managed to maximize *political* rather than social objectives.

Regulatory restrictions

An additional route of exploration should focus on the impact of regulatory restrictions on banks' activities. For example, in some countries banks have historically been authorized to own and control nonfinancial firms; and nonfinancial companies have been able to hold equity positions in commercial banks. In addition, banks have been able to operate in other markets through insurance underwriting and selling or through the underwriting and brokering of securities. The economic role of banking market power may be affected by the regulatory environment in which banks operate. For example, the mechanism proposed by Petersen and Rajan (1995) through which market power is needed for banks to establish lending relationships assumes that banks fund firms with traditional debt rather than equity finance. If a bank were authorized to finance via equity, the bank would participate in future profit sharing regardless of whether the firm maintains a lending relationship.

Therefore, it is possible that competitive banks allowed to provide equity finance would have the incentive to establish lending relationships. In such a world, the positive effect of bank concentration for firm growth identified in the empirical analysis described earlier may be less important.[5]

Moreover, in an environment where banks are authorized to operate in multiple markets (such as securities, real estate, and insurance), one could argue that, facing cross-market competitive pressures, banks in concentrated markets may be less able to extract rents. Therefore, in economies where banks are less restricted in their activities, the negative effects of bank market power may be of lower magnitude. This line of study is all the more relevant for the U.S. in light of the recent passage of the Financial Services Modernization Act.

Cetorelli and Gambera (2001) look at this issue, using a control variable that ranks countries according to how restrictive is the regulatory environment for banks.[6] However, they do not find significant evidence that the regulatory environment affects the role of banking market structure. More research along this line of inquiry is in order.

Bank concentration and concentration in other sectors

Does bank concentration "transmit" to other industries? In other words, does a concentrated banking sector lead to the formation of concentrated industry sectors, with fewer and larger firms? The effect on the market structure of industry sectors represents a novel dimension of analysis of the economic role of banking market structure. If the evidence in Cetorelli and Gambera suggests that bank concentration may help spur growth by favoring entry of young firms, could it still be the case that over time the concentrated banking industry leads to the emergence of concentration of ownership and control in those sectors that the banks helped to grow?

What determines industries' market structure? There is a literature in corporate finance focusing on the determinants of firms' size, in most cases the best available measure of an industry's market structure. If, all else equal, a sector is formed by a few, large firms, then that sector is relatively more concentrated, while if the same sector is formed by a relatively large number of smaller firms, then the sector is relatively unconcentrated. In a quite exhaustive work, Kumar, Rajan, and Zingales (1999) mention a large number of determinants of firm size and test their empirical significance. Evaluating industry-specific factors, the authors suggest that capital-intensive industries, industries with higher wages, and those with higher R&D (research and development) intensity exhibit larger firms. Looking at country (market) factors that are common across industries, countries with a better judicial system and those with higher human capital have industries with larger firms.

How does banking market structure fit within the theories of industry market structure determinants? Theories of industrial organization suggest that barriers to entry shape market structure. To the extent that banking market structure affects the availability of external finance, it acts as a barrier to entry. However, whether increasing bank concentration leads to more or less concentration in industry sectors, that is, whether it imposes a higher or lower barrier to entry, is a priori ambiguous. On the one hand, one could argue, according to the empirical evidence shown above, that a more "monopolistic" bank may enhance the growth of firms in earlier stages. Later on, as the sector matures, it may favor lending to the now incumbent firms over potential new entrants, a rationale that would be consistent with Petersen and Rajan (1995). In fact, what drove their monopolistic bank to finance the young firms in the first place was the opportunity to "participate," via rent extraction, in the future stream of profits when firms became established. The entry of new firms at more mature stages, by increasing market competition, would undermine the profitability of the incumbents. Hence, the bank might have an incentive to constrain the access to credit of new firms in more mature sectors. A second, separate argument would maintain that managers of banks in concentrated markets might have very close relationships with incumbent clients (for example, through membership of client companies' boards of directors and resulting participation in their management) and might be led by strategic decisions, not necessarily related to the bank's own profit maximization, to support these incumbents at the expense of prospective entrants. Either argument, therefore, suggests that bank concentration should lead to increasing concentration in industry sectors.

On the other hand, one could also argue that banks' ultimate goal of profit maximization could lead banks to continuously favor new entrants that, endowed with higher return projects and more innovative technologies, would guarantee higher bank profits. In this case, bank concentration should preserve unconcentrated industries, not contribute to the formation of large firms with significant market power.

Cetorelli (2001) analyzes this issue, using a data set comprising 35 manufacturing industries from 17 OECD countries. He finds that the average size of

firms in sectors highly dependent on external finance is indeed *larger* in countries with a more concentrated banking industry. Following Rajan and Zingales (1998) and Cetorelli and Gambera (2001), Cetorelli (2001) exploits industry variation along the dimension of external financial dependence to establish the empirical result: Whether bank concentration has a positive or negative effect on industry concentration, the effect should be especially strong on sectors that are relatively more dependent on bank finance. Therefore, Cetorelli (2001) examines whether industry concentration in sectors highly dependent on external finance is disproportionately higher or lower in countries whose banking market is more concentrated. The study makes a more sophisticated use of sector-specific information. Since the theoretical underpinnings suggest that bank concentration may play a role in industry market structure by favoring or not favoring clients with whom the banks already have long-term relationships (industry incumbents), one would expect to see an effect in those industry sectors whose *mature* firms are more dependent on external finance. If the effect is negative, it would suggest that even in sectors where mature firms are especially dependent on external finance, banks still allow entry of new firms, thus reducing the concentration of market shares among old incumbents. If the effect is positive, this would suggest that bank concentration enhances concentration in industry sectors.

A qualitative analysis of the effect of bank concentration on firm size is presented in tables 1 and 2. Table 1 reports mean values of average firm size, calculated for the data set of 35 manufacturing sectors in the 17 countries used by Cetorelli (2001). The measure of average firm size is the ratio of total value added of sector *j* in country *k* with the total number of establishments in the same sector and the same country. "Low" and "high" dependence refers to sectors, respectively, below and above the median in the distribution of external financial dependence. Similarly, low and high bank concentration refers to countries with a level of bank concentration, respectively, below and above the median of the cross-country distribution. Therefore, the table indicates, for example, that the mean of firm size of low dependent sectors in countries with relatively low bank concentration is $24.59 million, while the mean of firm size of the same sectors in countries characterized by high bank concentration is $12.3 million.

The numbers in the table allow me to make three main observations. First, low bank concentration countries have firms of larger size across all sectors (24.59 > 12.30 and 6.75 > 5.16). This indiscriminate

TABLE 1

Firm size of high- and low-dependence sectors in high and low bank concentration countries

	Low bank concentration	High bank concentration
	(- - - - - dollars in millions - - - - -)	
Low external dependence	+24.59	+12.30
High external dependence	+6.75	+5.16

Notes: Low external financial dependence sectors are below the median of the external financial dependence distribution. High external financial dependence sectors are above the median of the external financial dependence distribution. Similarly, low bank concentration countries have a bank concentration measure below the median, while high concentration countries have a concentration measure above the median. The numbers in the table are mean values across sectors of average firm size for each of the four clusters.

effect of bank concentration, rather than being due to the specific functioning of the banking market, is likely to be the result of a country *fixed effect*, that is, a characteristic common across all industries in the same country that affects *both* bank concentration and firm size. In particular, bank concentration is typically inversely related to the size of a country, as proxied by total employment, total population, or total income. Larger countries, in other words, have smaller bank concentration. The numbers in the table indicate that firms are larger in larger countries.[7]

Second, low-dependence sectors have larger firms across all countries, regardless of bank concentration (24.59 > 6.75 and 12.30 > 5.16). This effect, which is confirmed by Kumar, Rajan, and Zingales (1999), is likely to be due to industry fixed effects, that is, industry-specific technological factors that carry across countries. There is not an obvious prior to explain why sectors that are more dependent on external finance should have smaller firms. One possibility is that this result may indicate the indirect effect of financial constraints on firm size, assuming that there is not a strong correlation between bank concentration and the extent of financial constraints: Harder access to sources of finance should restrict the growth of existing firms, and this should particularly affect the sectors that rely more heavily on external sources of funding.[8]

My third observation is about the specific effect of bank concentration. As I mentioned earlier, whatever the effect of bank concentration on firm size, I expect it should be especially strong in sectors that are highly dependent on external finance. Although low-dependence sectors have larger firms, notice that

in countries with lower bank concentration, firms in low-dependence sectors are about four times as large as those in high-dependence sectors $\left(\frac{24.59}{6.75} = 3.64\right)$. However, in countries characterized by high bank concentration, firms in low-dependence sectors are only 2.5 times as large $\left(\frac{12.30}{5.16} = 2.4\right)$. These numbers are consistent with the argument that bank concentration contributes to increased firm size in industry sectors that are more bank dependent, relative to less dependent sectors.

I refine the analysis based on table 1 by attempting to purge the measure of firm size by industry- and country-specific factors. First, I regress average firm size on industry and country dummy variables. The series of residuals of the regression is a "cleaner" measure of sectoral firm size. The resulting numbers indicating firm size will be either positive or negative. A positive number shows that a certain sector in a certain country has firm size in excess of what its industry and country factors would indicate. Vice versa, negative numbers indicate sectors with firm size smaller than is accounted for by industry and country factors. Table 2 reports the mean value of the residuals for sectors below and above the median in external financial dependence in countries below and above the median in bank concentration. The first two observations I made regarding table 1 do not apply here, since any industry- or country-specific effect has been flushed out. In particular, it is no longer true that firms in countries with low bank concentration are larger regardless of the level of external dependence and that firms in low-dependence sectors are larger regardless of the level of bank concentration in a country. What about the specific effect of bank concentration? Note that in a country with low bank concentration, firms in low-dependence sectors are larger. Such firms have positive residuals on average, while high-dependence sector firms in the same countries have negative residuals. However, the pattern is exactly reversed when we move to countries with high bank concentration. In other words, we might formulate the following artificial experiment: What happens to firm size across sectors if a country increases bank concentration? The firm size of the most dependent sectors (again, those most affected by banks) goes from being below average to well above average. This confirms a positive effect of bank concentration on the firm size of highly dependent sectors.

Of course, these results are only suggestive. More convincing evidence would require a full-fledged econometric analysis, taking into account possible alternative explanations of this finding and testing for robustness. A complete investigation is found in Cetorelli (2001).

TABLE 2

Residual firm size net of industry and country fixed effects

	Low bank concentration	High bank concentration
Low external dependence	+2.69	–2.99
High external dependence	–2.47	+2.78

Notes: Low and high external financial dependence sectors and low and high bank concentration countries are as defined in table 1. The numbers in the table are mean values, calculated for each of the four clusters, of the residuals of a regression of average firm size on industry and country dummies.

Conclusion

This article presents an overview of the latest research on the economic role of bank competition. Contrary to the received wisdom that competition in the banking industry is necessarily welfare-enhancing, recent research has identified possible channels through which bank competition may generate negative economic effects. The main conclusion from the reading of the current body of research is that neither extreme—monopoly or perfect competition—may be the most desirable market structure for the banking sector. In advocating policies affecting the degree of bank competition, the regulator faces a tradeoff. While more competition is likely to lead to a larger *quantity* of credit, more market power should increase banks' incentives to produce information on prospective borrowers, thus leading to a higher *quality* of the applicant pool. Another related lesson to learn from this intellectual debate is that, in analyzing the role of bank competition, we should not restrict the investigation to its impact on the credit market, but rather support a broader approach which takes into account the fact that specific characteristics of the banking industry, such as its market structure, also affect various dimensions of other sectors of the economy. Examples of such interactions are the heterogenous effect on the growth potential and market structure of other industry sectors. Hence, the regulation of the banking industry has potentially important effects on the conduct of firms in other industries. For example, banking market structure may affect pricing strategies and incentives to innovation in other industry sectors.

NOTES

[1]The title of this article is taken from that of a conference cosponsored by The Wharton School of the University of Pennsylvania and by the Centre for Financial Studies of the Goethe University in Frankfurt, held in Frankfurt, Germany on April 7–8, 2000. The main goal of the conference, in the words of the organizers, was to develop a debate on whether bank competition should be seen as socially desirable. The conference program and papers are available on the Internet at http://fic.wharton.upenn.edu/fic/wfic/frankfurt2000.html.

[2]As recognized by Bhattacharya and Thakor (1993), "bank loans are special in that they signal quality in a way that other forms of credit do not" (p. 3).

[3]Fischer (2000) provides evidence from German data that, in more concentrated markets, banks produce more information in their lending activity.

[4]Note that in this analysis I have not even mentioned the potential effect of banking market structure on systemic risk and overall financial fragility. Hellman, Murdock, and Stiglitz (2000), for example, show theoretically that increases in competition, as determined by financial market liberalization, lower profits. Lower profits reduce banks' franchise value, and lower franchise value encourages banks to take more risk.

[5]In this respect, one can interpret market power as an implicit equity stake that the bank has in the firm it is financing.

[6]This cross-country indicator was put together by Barth, Caprio, and Levine (2000).

[7]The simple correlation between firm size and total income in the data set used in Cetorelli (2001) is +0.13 and highly significant. The correlation between bank concentration and total income is –0.73, also highly significant.

[8]But it could also work in the other direction: Financial constraints may impede entry of prospective new firms. The argument above would suggest that the first effect dominates the second one (this argument is also in Kumar, Rajan, and Zingales, 1999).

REFERENCES

Barth, J., G. Caprio, and R. Levine, 2000, "Banking systems around the world: Do regulation and ownership affect performance and stability?," in *Prudential Supervision: What Works and What Doesn't*, Frederic Mishkin (ed.), Cambridge, MA: National Bureau of Economic Research Press.

Bhattacharya, S., and A. V. Thakor, 1993, "Contemporary banking theory," *Journal of Financial Intermediation*, Vol. 3, pp. 2–50.

Black, S., and P. Strahan, 2000, "Entrepreneurship and the structure of the banking industry," mimeo.

Bonaccorsi Di Patti, E., and G. Dell'Ariccia, 2000, "Bank competition and firm creation," mimeo.

Cao, M., and S. Shi, 2000, "Screening, bidding, and the loan market," Queen's University, mimeo.

Cetorelli, N., 2001, "Does bank concentration lead to concentration in industrial sectors?," Federal Reserve Bank of Chicago, working paper, No. WP-2001-01.

_____, 1997, "The role of credit market competition on lending strategies and on capital accumulation," Federal Reserve Bank of Chicago, working paper, No. WP-97-14.

Cetorelli, N., and M. Gambera, 2001, "Banking market structure, financial dependence, and growth: International evidence from industry data," *Journal of Finance*, Vol. 56, No. 2, pp. 617–648.

Cetorelli, Nicola, and Pietro F. Peretto, 2000, "Oligopoly banking and capital accumulation," Federal Reserve Bank of Chicago, working paper, No. 2000-12.

Collender, R. N., and S. Shaffer, 2000, "Local bank office ownership, deposit control, market structure, and economic growth," U.S. Department of Agriculture, technical bulletin, No. 1886.

Dell'Ariccia, G., 2000, "Learning by lending, competition, and screening incentives in the banking industry," International Monetary Fund, mimeo.

Fischer, K. H., 2000, "Acquisition of information in loan markets and bank market power. An empirical investigation," Goethe University Frankfurt, mimeo.

Guzman, M., 2000, "Bank structure, capital accumulation, and growth: A simple macroeconomic model," *Economic Theory*, Vol. 16, No. 2, pp. 421–455.

Hellman, T., K. Murdock, and J. Stiglitz, 2000, "Liberalization, moral hazard in banking, and prudential regulation: Are capital requirements enough?," *American Economic Review*, Vol. 90, pp. 147–165.

Jayaratne, J., and P. Strahan, 1996, "The finance-growth nexus: Evidence from bank branch deregulation," *Quarterly Journal of Economics*, Vol. 111, pp. 639–670.

Kumar, K., R. Rajan, and L. Zingales, 1999, "What determines firm size?," University of Chicago, working paper.

La Porta, R., F. Lopez-de-Silanes, and A. Schleifer, 2000, "Government ownership of banks," Harvard University, mimeo.

Manove, M., J. Padilla, and M. Pagano, 2000, "Collateral vs. project screening: A model of lazy banks," Centre for Economic Policy Research, discussion paper, No. 2439.

Pagano, M., 1993, "Financial markets and growth. An overview," *European Economic Review*, Vol. 37, pp. 613–622.

Petersen, M. A., and R. G. Rajan, 1995, "The effect of credit market competition on lending relationship," *Quarterly Journal of Economics*, Vol. 110, pp. 407–443.

Rajan, R. G., and L. Zingales, 1998, "Financial dependence and growth," *American Economic Review*, Vol. 88, No. 3, pp. 559–586.

Shaffer, S., 1998, "The winner's curse in banking," *Journal of Financial Intermediation*, Vol. 7, No. 4, pp. 359–392.

Vittas, D., (ed.), 1992, *Financial Regulation. Changing the Rules of the Games*, Washington, DC: World Bank.