The Causes and Effects of Financial Modernization

Based on a speech given by President Santomero to the Pennsylvania Bankers Association, Charleston, SC, May 7, 2001

BY ANTHONY M. SANTOMERO

P assage of the Gramm-Leach-Bliley (GLB) Act, also known as the Financial Services Modernization Act, is just the latest step in the "relentless process of eroding the constraints placed on the financial marketplace during the Great Depression." So says President Santomero in this issue's *Third Dimension*. In addition, President Santomero sketches the history of financial services law since the 1930s. He then looks at the impact that GLB — especially its creation of financial holding companies — has had and will continue to have on the financial services industry.

In reality, "financial modernization" is not an event or a law; it is the dominant theme of the past 50 years of American finance. It signifies the erosion of arbitrary constraints that have divided the financial marketplace since the Great Depression. Therefore, describing the causes of financial modernization requires beginning then.

The Glass-Steagall Act of 1933 was enacted to protect consumers and the economy from the conflict of interest that, conventional wisdom held, contributed to the Great Depression. By separating deposit-taking activity from the underwriting of securities, the Glass-Steagall Act created a highly regimented financial services landscape. Commercial banks were limited to lending and deposit gathering. Thrifts were mortgage lenders. Investment banks served as underwriters and brokers of both stocks and bonds. And insurance firms had the profitable niche of actuarial products. Additional constraints were geographic in nature. Congress left in place a framework that encouraged state prohibitions on bank branching, leaving county and state borders as geographical boundaries on banks.

Congress should have anticipated the deterioration of the neat pigeonholes to which the financial industry was relegated. While useful in augmenting consumer confidence during the Depression, the boundaries became increasingly anachronistic in post-war America. Market pressure to expand product offerings and consumer desire to better meet financial needs, coupled with legal ingenuity and effective lobbying, were too powerful to allow these market constraints to survive indefinitely. Supplemented with the capabilities of computers and telecommunication, the evolutionary pace of financial-sector convergence accelerated greatly. By the 1970s, the very nature of banking had been changed forever.

In corporate finance, large, stable firms like General Motors and General Electric had long been the banking industry's best customers. But by the 1970s, many corporations found borrowing from banks to be less efficient than issuing direct capital market obligations. Bond traders could use computer technology to assess the merits of noninvestment-grade bonds, and they saw their industry boom at the expense of bankers. Innovative nonfinancial firms developed their own capacity to finance consumer debt by directly tapping the capital market, and they cut banks out of the loop.

At the same time, consumers no longer saw their traditional local bank as the only option for their savings
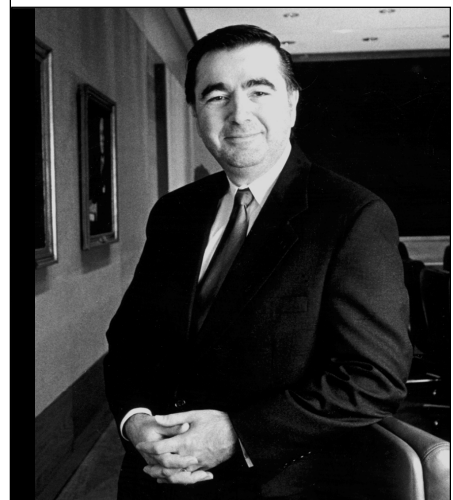
Anthony M. Santomero, President, Federal Reserve Bank of Philadelphia

balances. While they generally relied on a community bank or thrift for home mortgage loans, many consumers sought better returns for deposits through more sophisticated instruments. What was formerly deposited in a checking or savings account was now likely to be invested in a money market mutual fund or a cash management account or directly into securities. The money market mutual fund industry, which could not exist prior to computerization, held billions of dollars by the 1970s.

Traditional lenders, witnessing the drop in corporate and consumer deposits as well as loan demand, were eager to offer new products and find new sources of revenue. Technology did empower commercial banks to offer some new products and conveniences to their customers, such as the expanded use of credit cards, ATMs, and phone banking. But government often blocked their ability to compete within their traditional customer bases. Regulation Q, for example, forbade banks from offering competitive rates on checking accounts. Trying to stay competitive, many banks offered a completely new banking product — the toaster — as an incentive to open an account.

Such obstacles left bankers demanding relief through relaxed regulation, entry into new markets, and the ability to expand more freely across state borders. The government's response was to give them all three.

Action began at the state level when Maine enacted legislation permitting out-of-state entry. At the national level, Congress allowed banks to offer more competitive interest rates on deposits in 1980, ending the ill-conceived era of toaster banking. The Garn-St. Germain Act of 1982 allowed banks to cross state boundaries to acquire troubled banks. The Federal Reserve permitted bank holding companies to acquire discount securities brokers in 1983. In 1987, the Fed blessed limited securities underwriting under

the bank holding company umbrella — then expanded the limits in 1989 and again in 1996. The Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994 removed constraints on bank holding company acquisitions across state lines and also permitted banks to branch interstate if

> **Traditional lenders, witnessing the drop in corporate and consumer deposits as well as loan demand, were eager to offer new products and find new sources of revenue.**

permitted by state law. Interstate and regional banking had begun in earnest.

By the mid-1990s, the process of evolutionary convergence had transformed the financial services landscape. Commercial banks were brokering insurance and underwriting securities subject to percentage caps. Insurance companies, many of which had merged with investment banks, offered new risk-management products with all the characteristics of securities. Home mortgages were packaged into securities. Thrifts, credit unions, and commercial banks offered similar consumer products to their members. The money market provided more efficient transfers of capital. Major commercial firms had their own finance companies or even a thrift. And with mergers and acquisitions, the size of financial conglomerates swelled to unprecedented new levels.

These developments made economic sense. In many cases they were the only rational courses of action that could be taken by Congress, the Federal Reserve, the Federal Deposit Insurance Corporation, the Office of the Comptroller of the Currency, or individual states. But these actions stretched the credibility of the rules. Often, the rulings of bank regulators seemed like reversals of established policy, because bank products emerged

despite regulatory prohibitions or regardless of precedent. Now terms such as "nonbank banks" and "the facilitation of commercial paper placement" entered the lexicon. And as complexity rose, smaller institutions found themselves at a competitive disadvantage. By the mid-1990s, large

sections of federal banking law resembled relics of a bygone era.

The contrast between the inadequacy of existing legislation and the reality of a new financial services paradigm was made clear in April 1998 when Citicorp and Travelers Group proposed a $70 billion merger. The creation of Citigroup — America's largest financial conglomerate, with businesses ranging from banking to insurance to securities underwriting — demonstrated the inadequacy of the legislative and regulatory patches of the previous 20 years. Congress knew it had to stop debating financial laws and respond. Within a year, both the House and the Senate had passed legislation to bring our financial laws into the modern age. With President Clinton's signature in November 1999, the Gramm-Leach-Bliley (GLB) Act, also known as the Financial Services Modernization Act, became law.

GLB provides a unified legal framework that standardizes financial convergence. Its centerpiece is the creation of an entity called a financial holding company, or FHC. Once a financial organization obtains the FHC designation, it can house a complete family of financial activities through distinct affiliates. Each affiliate is still overseen by its traditional functional regulator. The Federal Reserve

continues to oversee the FHC, much as it oversees all the bank holding companies, or BHCs, of both yesterday and today.

However, while GLB established a new legal framework for financial convergence, it did not change the underlying realities driving the marketplace. Technology, demographics, and customer needs are the forces that have determined and will continue to determine the structure of the financial services industry.

But while GLB will not change the nature of the industry, it will bring the financial services industry to convergence in a more expeditious and orderly manner — but one that still holds a few surprises. For example, before GLB was enacted, some predicted that many banks and other financial service organizations would quickly seek FHC status and begin offering "one-stop shopping" for financial services to their target customers. It's been about 18 months since organizations could apply to become FHCs. Thus far, things have not turned out as predicted.

As of August of this year, less than 20 percent of top-tier bank holding companies had converted to an FHC. The percentage of investment banks, brokerage houses, and insurance companies that converted is much smaller.

Not surprisingly, the largest multi-product institutions have led the way. Before GLB, these large organizations were constrained from pursuing a "financial supermarket" strategy, so they acted swiftly to maximize that opportunity.

A number of relatively small banks and small bank holding companies also have found reason to obtain FHC status. Indeed, fully two-thirds of current FHCs have assets of less than $500 million. These institutions sought this status not because they have immediate plans to expand their product

offerings, but because the designation presented a relatively low-cost option for future expansion. In general, these local or regional BHCs have less complex corporate structures and are well-capitalized. So, getting the

A number of relatively small banks and small bank holding companies also have found reason to obtain FHC status.

designation proved relatively easy, and these institutions will be prepared for good future opportunities.

Nonetheless, only a small percentage of the total number of firms many suspected would be eager to benefit from the new law have chosen to seek the designation. Why have so few financial firms elected to become FHCs? Why has the pace of cross-industry acquisition been so slow? Undoubtedly, there are many reasons why more financial institutions have not rushed to obtain a designation that allegedly allows them to be all things to all customers. However, one seems particularly relevant.

Perhaps I am too much of an economist, but I believe that many institutions have done a simple calculation. They have already adapted to BHC structure. They have been successful in delivering financial services to their market area through a combination of bank and

nonbank subsidiaries, coupled with the increasing use of strategic alliances and outsourcing. Their operating structures are in place and have been effective.

By contrast, I believe that many of these institutions see no immediate benefits of converting to an FHC and remain uncertain as to the longer term implications of FHC status.

Over time, the potential benefits of the FHC structure will be clarified by developments both in the marketplace and in regulatory pronouncements. Circumstances will illustrate whether the added flexibility afforded institutions operating under an FHC charter offers additional, exclusive profit opportunities. Meanwhile, regulatory policies and procedures will reveal the parameters under which FHCs must operate.

A number of the detailed regulations necessary to implement Gramm-Leach-Bliley have yet to be offered for public comment by the Fed, and none of the law's provisions have undergone "trial by fire." Under-

standably, the rules that the Federal Reserve has offered for comment are being scrutinized for indications of the Fed's intent and its appreciation of industry conditions.

This was made abundantly clear by the reaction to the proposed rule that bank holding companies' merchant banking activity should be subject to a 50 percent capital charge. The comment period worked as intended — and the Federal Reserve substantially altered the rule — but the

proposed in recent years. Comments were due August 15, and the input of affected organizations will be constructive and useful in determining the final form of this landmark regulation.

I also believe that the industry is interested in how relationships among regulatory agencies will unfold in this new environment. The Federal Reserve's new role as umbrella supervisor of financial holding companies is similar to its role in supervising bank holding

marketplace during the Great Depression. With the passage of the Financial Modernization Act of 1999 and the implementation of the FHC structure, that process took a big step forward.

What will the financial services industry look like in the future? It is hard to say, but there is some agreement — at least in broad strokes. There will surely be a handful of financial behemoths offering one-stop shopping to businesses and consumers. Their outlines and their names seem to be emerging daily.

Beyond these few that will attempt to be all things to all people, a large number of institutions will remain. These may be described as niche players, which will choose to concentrate on either a geographic area or a product set. In their chosen market segment, they will remain credible, even fierce competitors. Single-product providers, such as credit card and mortgage servicing companies, will remain. Community banks will still be effective competitors, both in markets for small-business lending and personalized consumer service. These smaller banks are quick to adjust to changes in customer needs, and they will be able to compete effectively as well.

> **In short, the future holds more innovation for firms of all sizes. The needs of customers...will continue to evolve, and financial service providers will, as always, adapt to meet their needs.**

episode undoubtedly left some lingering apprehension.

Another important step toward implementing Gramm-Leach-Bliley was taken in early May when the Board of Governors announced that it was seeking comment on the long-awaited Regulation W. This proposed rule seeks to implement section 23 A and B of the Federal Reserve Act and to define permissible transactions between a bank and its affiliates. In the post financial modernization world, bank affiliations can and do extend to many kinds of institutions. Protecting insured deposits from improper transfer to an affiliate is vital to the safety and soundness of our national economy and one of the key functions of this regulation.

Following enactment of Gramm-Leach-Bliley, implementing regulations for section 23 A and B became more possible and more necessary. But despite a basis in precedent, Regulation W is as complex and detailed as any federal regulation

companies. However, the Fed's future success entails increased communication, cooperation, and coordination with the many supervisors of the more-diversified financial holding companies. As the Fed begins redefining its working relationship with other regulators, it will answer many of the questions of importance to securities and insurance-based firms.

As we develop the rules and refine the regulators' roles in the financial holding company, I believe that FHCs will emerge as entities with the flexibility and functionality to meet the demands of the marketplace without unnecessary or onerous regulatory burden. As this becomes clear, I expect the number of financial firms electing to establish financial holding companies will increase.

At the start of this article, I said that financial modernization is not a single event or law, but rather a relentless process of eroding the constraints placed on the financial

In short, the future holds more innovation for firms of all sizes. The needs of customers, be they individuals or organizations, will continue to evolve, and financial service providers will, as always, adapt to meet their needs. Gramm-Leach-Bliley recognizes that this is the nature of the marketplace. Those who find ways to seize the opportunities this law offers will benefit in the financial marketplace of the future and will be the first to reap the rewards. **ⓑⓡ**

Privacy Matters:
Payment Cards Center Workshop on the Right to Privacy and the Financial Services Industry

BY SARAH A. BURKE

P assage of the Gramm-Leach-Bliley (GLB) Act in 1999 re-opened the debate on consumers' right to privacy in financial transactions. To broaden awareness of this debate, the Philadelphia Fed's Payment Cards Center sponsored a workshop, led by University of Pennsylvania law professor Anita L. Allen. Professor Allen opened the meeting with a general discussion of privacy issues, then focused on privacy provisions of GLB. In this article, Sally Burke outlines some of the primary concerns and summarizes Professor Allen's presentation.

Also known as the Financial Services Modernization Act, Gramm-Leach-Bliley (GLB) allows financial institutions to engage in certain types of activities that were formerly prohibited.

In effect, GLB repealed sections 20 and 32 of the Glass-Steagall Act, which, among other things, separated commercial and investment banking. GLB also created an entity called a financial holding company (FHC). Any bank holding company that qualifies to be an FHC may engage in a broad range of finance-related activities, including underwriting insurance and securities. This closer union between banks and other financial services organizations increased concerns about how customer information gathered by financial institutions would be shared, especially with unaffiliated third parties.

The privacy provisions of GLB describe the conditions under which financial institutions[1] may disclose nonpublic personal information about consumers to nonaffiliated third parties, require such institutions to provide notice to their customers about their privacy policies, and permit the consumer to opt out of those disclosures, subject to certain exceptions. Congress has provided broad rule-making authority to eight federal agencies, each of which regulates a different aspect of the financial services industry.[2]

The agencies' privacy regulations apply to financial institutions only with respect to the nonpublic personal information about individuals who obtain financial products or services primarily for personal, family, or household purposes. The privacy regulations do not apply to information about companies or about individuals who obtain financial products or services for business, commercial, or agricultural purposes.

Earlier this year, the Payment Cards Center of the Federal Reserve Bank of Philadelphia sponsored a workshop with Anita L. Allen, a professor of law at the University of Pennsylvania. Professor Allen, who has written and lectured extensively about the legal aspects of privacy, led a discussion with Philadelphia Fed officers and staff about privacy issues in general and privacy provisions under GLB in particular. Her remarks provided a

Sally Burke is senior editor and publications manager in the Public Affairs Department of the Philadelphia Fed.

---

[1] For the purposes of the privacy provisions, the term "financial institution" is defined to mean any institution — whether or not affiliated with a bank — that engages in activities permissible for a financial holding company. Thus, the term would include banks, thrifts, mortgage companies, and insurance and securities firms.

---

[2] In accordance with the statutory mandate, the agencies, including the Board, worked together to implement privacy regulations that contain substantively identical provisions. The Board's privacy rule, Regulation P (12 C.F.R. Part 216), applies to the U.S. offices of entities for which the Board has primary supervisory authority.

historical timeline for these issues and a context for GLB.

To start, Professor Allen offered her definition of privacy: "modes by which people, personal information, certain personal property, and personal decision-making can be made less accessible to others." She noted further that privacy is protected not only by law but also "by cultural norms, ethics, and business and professional practices." She also listed four types of privacy: informational, physical, decisional, and proprietary. GLB privacy provisions fall mostly into the informational category. (See *Types of Privacy*.)

Of course, Professor Allen acknowledged that when we talk about privacy, a basic question arises: Why is it important? Because, Professor Allen stated, it involves factors such as personhood, individuality, personal and social relationships, autonomy, and tolerance, to name just a few. But, she cautioned, privacy rights are not absolute. Such rights must often be weighed against other considerations such as public health and national security. (See *Privacy vs. Other Values, Needs, and Policies*.)

The word "privacy" does not appear in the Constitution.

In fact, the word "privacy" does not appear in the Constitution; however, Professor Allen noted that the Supreme Court has interpreted five of the 10 original Bill of Rights guarantees and the 14th Amendment as protective of privacy. For example, the Court has stated that the search and seizure protections of the Fourth Amendment relate not only to the physical privacy of a citizen's home but also to the informational privacy of a citizen's papers, correspondence, conversations, and electronic communications.

Professor Allen believes that mistaken ideas about citizens' rights to privacy are quite common. That's one reason she thinks people don't shop around for another bank even when they're concerned about privacy – they *assume* that their depository institution protects their privacy as a matter of course.

Articles in the popular press support this belief that people have exaggerated notions about their right to privacy. In the March 2001 issue of *The Atlantic Monthly*, author Toby Lester states that people tend to assume that privacy "is one of the bedrock rights upon which American society is built." But as Lester's article, "The Reinvention of Privacy," points out, Americans originally thought of privacy as "a physical concept." Citing the work of Robert Ellis Smith, Lester says that for most Americans before the end of the 19th century, protecting one's privacy or

Types of Privacy	Examples
Informational Privacy (most important for GLB)	Informational privacy is at issue in cases about access to medical records, employer access to email, on-line anonymity, data encryption, and executive privilege. Confidentiality and secrecy are informational privacy concerns. Concerns about informational privacy go by many names, including secrecy, confidentiality, anonymity, security, data protection, and fair information practices.
Physical Privacy	Physical privacy is at issue in cases about government search and seizure, peeping toms, and "ambush" journalism. Seclusion and solitude are physical privacy concerns. The home is the traditional seat of physical privacy. Bodily integrity is sometimes an important physical privacy concern.
Decisional Privacy	Decisional privacy is at issue in cases about abortion rights and the right to assisted suicide. The rights of homosexuals and families to direct their own lives are commonly styled as privacy concerns in the decisional sense.
Proprietary Privacy	Proprietary privacy is at issue in cases about publicity rights, identity, and the ownership of the body. The rights of celebrities and others to control the attributes of their personal identities are commonly styled as privacy concerns in the proprietary sense.
Source: Professor Anita L. Allen, University of Pennsylvania Law School	

acquiring more of it was simply a matter of moving west, where "there were fewer people likely to know or care what one was doing." Today, although people still retain a sense of physical privacy about their homes and other property, privacy has acquired an abstract aspect as well, thanks to developments such as personal computers and the Internet.

However, PCs and cyberspace are just the most recent links in an age-old chain. In fact, technology has spurred interest in privacy issues before. Lester's article offers this example. In 1890, Samuel Warren and Louis Brandeis wrote an article called "The Right to Privacy" for the *Harvard Law Review*. Cameras and high-speed printing presses were the new technologies that prompted Warren and Brandeis to write their treatise.

Although issues about certain types of privacy have obviously been in the public consciousness for a long time, privacy as it relates to financial services is a relatively new phenomenon. Through the 1960s, Professor Allen said, financial services generally entailed a contractual relationship between consumers and their banks, and banks — as yet unhampered by legal considerations — had a lot of freedom to share information about customers.

But the 1960s saw a resurgence of interest in matters of privacy. Once again, technology drove the discussion. The development of computers in the 1960s led to concerns about how and where information was stored and who had access to it. The cold war and the domestic social and political movements of that decade also raised questions about surveillance, particularly government "spying" on private citizens.

Legislative action to address these concerns started to come about in the 1970s. In the financial services area, Congress passed the Fair Credit Reporting Act (FCRA) in October 1970. The FCRA, which applies only to consumers, covers the confidentiality,

accuracy, relevance, and proper use of credit information. This law also restricts access to consumers' credit reports. In 1974, the Privacy Act mandated "fair information practices" and limited third-party access to personal information contained in record systems. That same year, Congress passed the Freedom of Information Act (FOIA), which gave the public access to government records. But FOIA does contain exceptions for medical, personnel, and "similar files." The Right to Financial Privacy Act of 1978 extended the rights in these earlier laws by governing certain banking and financial transactions. Among other things, this act restrains the government's access to some types of financial information and prohibits the unauthorized release of records by financial institutions.

In the 1980s, Congress passed a string of legislation regarding a number of privacy issues; the most important for financial services was the Fair Credit and Charge Card Disclosure Act of 1988. This legislation expanded some of the disclosure provisions of the Truth in Lending Act. In short, it required all credit and charge card issuers to provide consumers with specific information on interest rates, fees, etc., in an easy-to-

read format or to provide a toll-free number and an address from which consumers could obtain such details.

Most recently, Congress passed GLB in 1999. Under its privacy provisions, GLB requires a financial institution to inform consumers that it may disclose – or reserve the right to disclose – "nonpublic personal information" to nonaffiliated third parties. In addition, consumers must be offered the opportunity to "opt out" of such disclosures, and the financial institution must give consumers "reasonable means" by which to exercise their opt-out right. The law further mandates that financial institutions must inform customers about information-sharing policies at the start of the relationship and annually thereafter. All financial services organizations had to comply with these provisions by July 1, 2001.

Of course, with the trend toward a global marketplace, a question arises concerning just how much protection consumers derive from the privacy provisions of GLB. Many national and international companies have so many affiliates that "nonpublic personal information" can legitimately be shared with numerous entities.

Privacy vs. Other Values, Needs, and Policies

Privacy vs.
- First Amendment Freedom of Speech and Press
- Newsworthiness of Information
- The Public's Right to Know About Government, Officials, and Businesses
- National Defense, Military Necessity
- Criminal Law Enforcement
- Public Health and Safety
- Employer Necessity or Business Profitability
- Government "Special Needs"
- Efficiency, Expense, or Administrative Necessity
- Fiduciary Values, e.g., Trust, Accountability, or Loyalty

Source: Professor Anita L. Allen, University of Pennsylvania Law School

who's watching whom, consumers must now worry about "the shift from Big Brother to Big Business."

Professor Allen also explained that GLB is an extension of the government intervention that began in the 1970s. Furthermore, it changes the relationship between banker and consumer by imposing a statutory obligation, effectively replacing the contractual relationship that previously existed.

The legal aspects of privacy, of course, have many more facets than those presented here. So, too, all of the details of the financial modernization legislation are beyond the scope of this article.[3] However, the Payment Cards

Center at the Federal Reserve Bank of Philadelphia hopes that the discussion with Anita Allen and the subject of privacy, especially as it relates to the financial services industry, will further stimulate consumers', regulators', and the industry's interest in this important topic.

As Peter Burns, director of the Center noted, "There is arguably no sector in financial services where the collection and management of consumer data are more central to the core business model than in the payment cards industry. Center-sponsored workshops and discussions with thoughtful observers such as Dr. Allen are important tools for helping to inform the underlying policy debate." ®

Moreover, GLB permits joint marketing arrangements with nonaffiliated third parties. Noting some of the social differences between today and 40 years ago, Dr. Allen, quoting sociologist Amitai Etzioni, stated that in matters of

---

[3] For an excellent summary of the provisions of Gramm-Leach-Bliley, see the Federal Reserve Bank of Philadelphia's *Banking Legislation and Policy* newsletter, Vol. 18, No. 4, October-December 1999. The article is available on the Bank's web site at: www.phil.frb.org/files/blp/blpq499.pdf. Or see the Philadelphia Fed's Supervision, Regulation, and Credit site at http://www.phil.frb.org/src/glba.html.

How Do Forecasts Respond to Changes in Monetary Policy?

BY LAURENCE BALL AND DEAN CROUSHORE

J ust as changes in atmospheric conditions affect weather forecasts, changes in monetary policy affect economic forecasts. When monetary policy shifts, forecasters change their predictions about growth and inflation. But does the economy change to the same extent that forecasts do? In this article, Laurence Ball and Dean Croushore examine forecasts from the *Survey of Professional Forecasters* to determine if forecasts and the economy respond in tandem or if there are significant differences.

Forecasts are important because they affect what people do. If the weather forecast calls for rain, people carry umbrellas and cancel outdoor activities. If the economic forecast calls for a rise in the unemployment rate, people will reduce their spending on consumer goods.

And just as atmospheric conditions affect weather forecasts, changes in monetary policy affect

Larry Ball is a professor of economics at Johns Hopkins University.

Dean Croushore (pictured left) is a vice president and economist in the Research Department of the Philadelphia Fed.

economic forecasts. If the Federal Reserve tightens monetary policy, forecasters predict slower economic growth and lower inflation; if the Fed eases monetary policy, forecasters predict faster growth and higher inflation. But does the economy change to the same extent the forecasts do?

To answer this question, we'll look at forecasts from a survey of professional economic forecasters. We'll see how the economy responds to a change in monetary policy compared with how forecasts respond, to determine if the responses are identical or if there are significant differences.[1]

Why should we care about

---

[1] This paper reports the results of the authors' joint research. For additional details of their research, see their working paper, cited in the References.

whether the economy changes to the same extent the forecasts do? If forecasts systematically respond differently than the economy does to a shift in monetary policy (that is, to a greater or lesser degree or with different timing), we might reach two conclusions: forecasters are irrational (since a good forecast should change in the same way the economy does) and forecasts aren't accurate guides to what happens in the economy when monetary policy changes.

Such conclusions can have repercussions. First, if forecasters are irrational, people will be less likely to believe their prognostications. Second, inaccurate forecasts may influence economic activity indirectly by setting up false expectations about how monetary policy will affect the economy. Acting on those expectations, people will behave in a certain way. But since people are misinformed about what effects monetary policy will have, they'll behave in a manner different from how they'd act if they had better information. Thus, monetary policy might affect real output in the economy partly because people were misinformed about its effects.

On the other hand, if forecasts align well with how the economy changes when monetary policy shifts, that's a sign that economic forecasters are rational. This alignment also eliminates the possibility that monetary policy affects the economy because people misinterpret its effects.

FORECAST DATA

To investigate forecasts, we'll use the *Survey of Professional Forecasters*

(SPF).[2] The survey, which began in 1968, reports the forecasts of economists throughout the business world, on Wall Street, and at consulting firms. The survey asks participants to provide their quarterly forecasts for 18 major macroeconomic variables, including real GDP and all of its components. The survey form typically runs four to six pages; sometimes the survey includes special questions, which vary depending on current economic conditions. Because of the amount of detail the survey asks for, economists who participate in the survey are those for whom forecasting represents a major part of their job responsibilities. The survey, which is run by the Federal Reserve Bank of Philadelphia, is the leading quarterly survey of U.S. economists' forecasts.[3]

How good are the forecasts overall? If we examine just the average across the forecasters in the survey, we'd like to know if that average forecast is reasonable. If you wanted a good forecast for future output growth or inflation, would these surveys be useful to you? The answer is yes. These surveys almost always pass analysts' statistical tests for accuracy. For example, Dean Croushore recently studied the inflation forecasts from the *Survey of Professional Forecasters* and several other surveys and found that the SPF forecasts were quite good, though there were periods in which SPF respondents made severe forecast errors. Those periods were most often

associated with oil-price shocks, mostly in the 1970s and early 1980s, when the economy performed poorly and inflation was rising dramatically.[4]

Figure 1 gives an overview of how accurate survey forecasts are. It shows the one-year-ahead forecasts for output growth made each quarter, from the fourth quarter of 1968 to the fourth quarter of 1999, compared with the data that show what actually happened. (For example, the forecast made in the fourth quarter of 1968 predicts output growth from the fourth quarter of 1968 to the fourth quarter of 1969. We compare the forecast with the actual data over the same period.) Figure 2 does the same for inflation forecasts. All the forecasts are looking one year ahead, and the date the forecast was

made is shown on the horizontal axis.

Figure 1 demonstrates that, for the most part, output forecasts are good, in the sense that, on average, the difference between the forecast and what actually happened was near zero.[5] Consequently, one-year-ahead output forecasts match up with the data fairly well. The forecasts aren't quite as volatile as the actual data, which is a characteristic of all good forecasts. But the general pattern of movement over time is the same for the two series. There have been no long periods in which forecasts were consistently too high or too low except, perhaps, in the late 1990s.

In Figure 2, you can see that inflation forecasts over the past 15 years were pretty good, but they were much

---

[4] For more details on these results, see Croushore's 1998 working paper and his 1996 *Business Review* article.

[5] For details, see the authors' 2001 working paper.

FIGURE 1

Mean Output Growth: Forecast and Actual
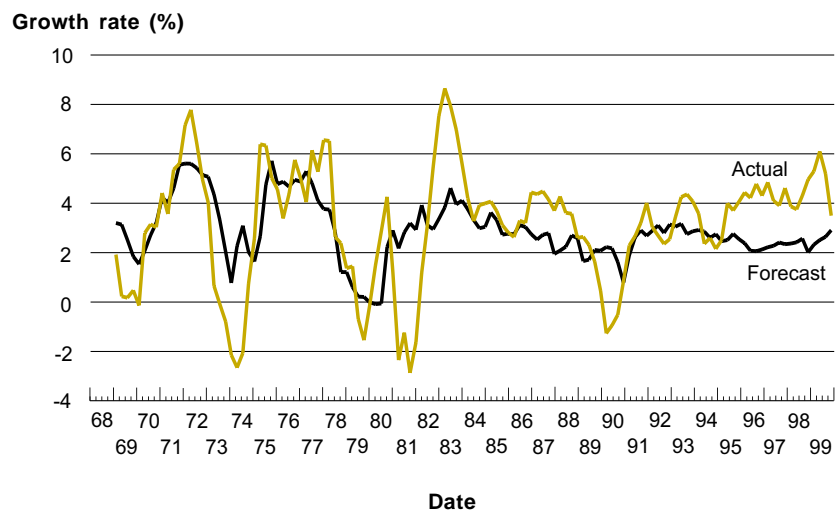
Growth rate (%)

Note: Dates shown are dates when one-year-ahead forecasts were made; actual is for one year ahead from date of forecast. For example, in 1968Q4, forecasters on average predicted that output growth would be 3.2% between 1968Q4 and 1969Q4; output growth turned out to be 1.9%.
Source: *Survey of Professional Forecasters* and authors' calculations.

worse in earlier years. In the late 1960s and throughout the 1970s, forecasts for inflation were too low, with errors averaging 1.6 percent; in the early 1980s, inflation forecasts were too high, with errors averaging 1.7 percent; and through much of the 1990s, inflation forecasts were again too high, but the errors were smaller, averaging 0.8 percent. Nonetheless, standard statistical tests suggest that, overall, the forecasts are not biased, that is, they weren't consistently wrong in one direction or another. Thus, they pass a simple test for accuracy.[6]

Measuring Monetary Policy. Given that the forecasts look fairly good overall, the question arises: how do the forecasts respond to changes in monetary policy? To answer that question, we need a quantitative measure of monetary policy. Economists often use a real interest rate, that is, the interest rate adjusted for expected inflation, as a variable for determining how monetary policy is changing. Since the Federal Reserve generally operates by targeting the federal funds rate, which is the interest rate on short-term loans between banks, our measure of monetary policy is the *real* federal funds rate.[7]

## COMPARING FORECASTS WITH REALITY

To see how well the forecasts compare with what actually happens in

[6] The inflation forecasts sometimes missed the mark, especially when there were big oil-price shocks, but they were not consistently wrong. For more on testing for bias in forecasts, see Croushore's 1996 article and his 1998 working paper.

[7] The real federal funds rate is defined as the nominal federal funds rate minus the expected inflation rate. Even if the survey's expected inflation rate turned out to be biased, the real federal funds rate defined this way would still be the correct measure of the stance of monetary policy because it's a key variable that people use in making economic decisions.

FIGURE 2

Mean Inflation: Forecast and Actual
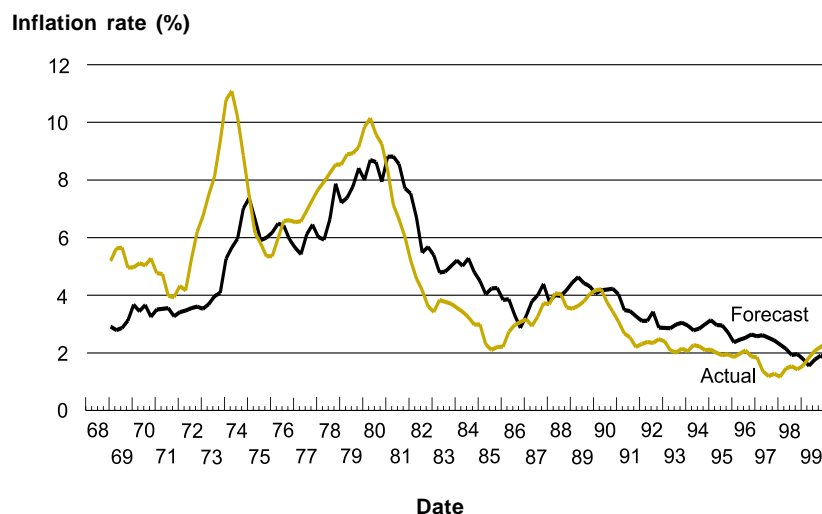
Note: Dates shown are dates when one-year-ahead forecasts were made; actual is for one year ahead from date of forecast. For example, in 1968Q4, forecasters on average predicted that inflation would be 2.9% between 1968Q4 and 1969Q4; inflation turned out to be 5.2%.
Source: *Survey of Professional Forecasters* and authors' calculations.

---

the economy, we'll break them into several parts. First, we'll look at a benchmark forecast formed using only past values of output or inflation, to get a rough idea of how output or inflation might change if there were no changes in monetary policy. Then, we'll compare each survey forecast with this benchmark forecast. Finally, we'll compare the survey forecast to what actually happened in the economy.

A Benchmark for Comparison. We're going to begin our analysis by using a simple model as a benchmark for comparison. A simple forecast of output growth is one based only on past data for real output growth.[8] Similarly, our benchmark model for inflation attempts to provide a useful forecast of inflation based solely on past inflation rates.

[8] In technical terms, this is called a univariate time-series model.

We chose this simple model as a benchmark because it ignores any past changes in monetary policy that are likely to affect output growth or inflation in the future. Then, by comparing the forecasts from this benchmark model with the forecasts made in our surveys, we can observe, in principle, how the survey forecasts respond to changes in monetary policy. Of course, if monetary policy doesn't change, the benchmark model's forecasts should be similar to the survey forecasts.

You might think that these types of models wouldn't be very good at forecasting; however, our tests suggest that they do very well. When we ran the forecasts through a battery of tests (see our working paper for details), they passed every one.

Measuring the Effects of Monetary Policy. To see how monetary policy affects output growth, we'll look

at the difference between actual output growth over the course of the year and our benchmark model's forecast for output growth over the same period. This difference is called the benchmark error. If monetary policy's effects on the economy are not fully reflected in the benchmark forecasts, we would expect to find that changes in monetary policy are associated with benchmark errors. In particular, we would think it likely that tighter monetary policy today (a higher real federal funds rate) would reduce future output growth but that our simple model wouldn't pick up this effect because the model doesn't incorporate information about monetary policy. So tighter monetary policy should be correlated with a negative value of the benchmark error. Similarly, easier monetary policy should be correlated with a positive benchmark error, since such policy would increase actual output growth but would not affect the benchmark forecast.

The simplest way to demonstrate this is a scatterplot showing the benchmark error, that is, the difference between actual output growth and the benchmark model's forecast on the vertical axis and the measure of monetary policy — in this case, the change in the real federal funds rate over the preceding year — on the horizontal axis (Figure 3a).[9] The plot

shows a clear negative relationship.[10] Tighter monetary policy, which is a positive change in the real federal funds rate, is associated with negative values of the benchmark error. A more formal statistical test confirms that the relationship is statistically significant.[11]

We also can examine differences between actual inflation and our benchmark forecast for inflation. In this case, tighter monetary policy is expected to lead to lower inflation than the univariate time-series model suggests. So increases in the real federal funds rate

would be correlated with negative values of the benchmark error. Similarly, declines in the real federal funds rate would be correlated with positive values.

When we look at the data on inflation and changes in monetary policy, we don't see a clear relationship, in part because monetary policy takes longer to act on inflation than on output. This suggests that we need to look at changes in monetary policy from longer ago. Indeed, if we look at the change in the real federal funds rate from two years to one year prior to the forecast, we see a negative impact, as expected, though the relationship is a bit weaker than in the case of output (Figure 3b). Again, statistical tests confirm this negative relationship.

Overall, tighter monetary policy reduces both future output and future inflation in a way that our benchmark forecasts do not pick up.

How Survey Forecasts Reflect Information About Monetary Policy. Next, let's examine how the survey forecasts reflect the fact that the economists surveyed make their forecasts using information about monetary policy. If they didn't use such

---

[9] To be consistent with the timing of the survey forecasts, and to ensure that the forecasters knew the value of the measure of monetary policy at the time their forecasts were made, we follow this timing convention: For a forecast made in a particular quarter (for example, the fourth quarter of 1968), we look at the change in the real federal funds rate over the year that ends in the previous quarter (for example, from the third quarter of 1967 to the third quarter of 1968). That's because the survey forecasts are made in the middle of the quarter, so the survey respondents don't yet know the value of the real federal funds rate for the quarter in which they make their forecasts (in this example, the fourth quarter of 1968).

[10] We've highlighted this relationship by drawing a regression line, which represents a linear relationship between the variables on the horizontal and vertical axes.

[11] Technically, we ran a regression of the forecast error on the one-year change in the real federal funds rate, then tested the significance of the regressor with an exclusion test. The test is slightly complicated by the fact that the observations are quarterly and the forecast horizon is one year ahead, so we must use a Newey-West procedure to adjust the covariance matrix. See our working paper for test details. We perform all tests at a 5 percent significance level.

information, we'd expect the survey forecast for output growth to be similar to that of our simple benchmark model. But if survey participants use information about monetary policy in setting their forecasts, the difference between the *survey forecast* and our *simple benchmark forecast* would vary depending on whether monetary policy was tight or easy. In particular, tighter

> Survey forecasts of output don't fall enough when monetary policy tightens, but survey forecasts of inflation decline by the right amount.

monetary policy (an increase in the real federal funds rate) would lead survey forecasts for output growth to be lower than our benchmark forecasts. That is, we'd expect the difference between these forecasts to be negative. Similarly, forecasters anticipating easier monetary policy (a decrease in the real federal funds rate) would expect growth to increase. Thus, survey forecasts would tend to be higher than the simple benchmark forecasts, so we'd expect the forecast difference to be positive. Again, the same type of analysis can be done for inflation as for output growth.

Let's repeat the analysis shown in Figure 3a, but this time we'll look at the difference between forecasts for output growth from the *Survey of Professional Forecasters* (SPF) and the benchmark forecasts. The same type of scatterplot shows a negative relationship (Figure 4a), which is what we expect. Tighter monetary policy (a positive value of the change in the real federal funds rate shown on the horizontal axis in the figure) is associated with a negative forecast difference. This

FIGURE 3a

The Effect of Monetary Policy on Output

Output growth over coming year
Actual minus benchmark forecast (%)
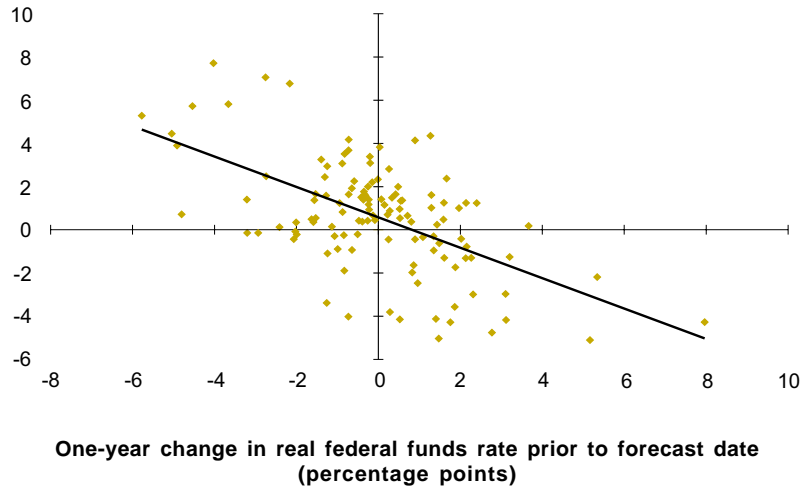
**One-year change in real federal funds rate prior to forecast date
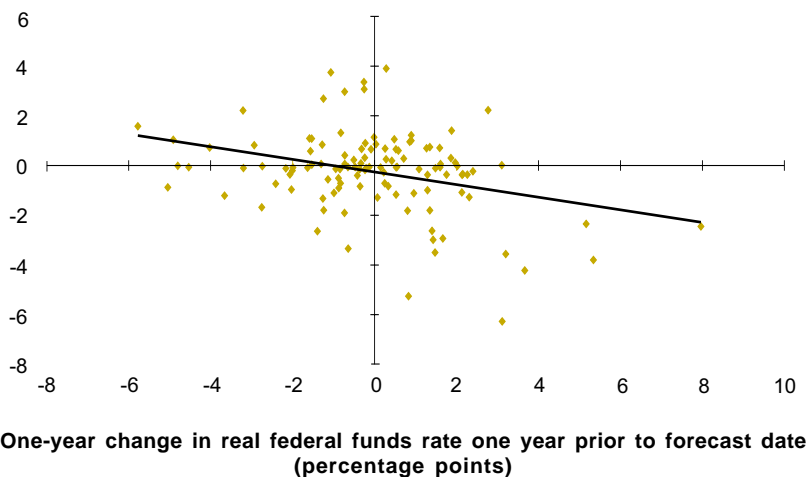(percentage points)**

Example: The difference between actual output growth and the benchmark forecast between 1981Q4 and 1982Q4 was -4.3 percentage points and the change in the real federal funds rate between 1980Q3 and 1981Q3 was 8.0 percentage points. This is the point farthest to the right in the figure. Note: A linear regression line is plotted.
Source: *Survey of Professional Forecasters* and authors' calculations.

FIGURE 3b

The Effect of Monetary Policy on Inflation

Inflation over coming year
Actual minus benchmark forecast (%)

**One-year change in real federal funds rate one year prior to forecast date
(percentage points)**

Example: The difference between actual inflation and the benchmark forecast between 1982Q4 and 1983Q4 was -2.4 percentage points and the change in the real federal funds rate between 1980Q3 and 1981Q3 was 8.0 percentage points. This is the point farthest to the right in the figure. Note: A linear regression line is plotted.
Source: *Survey of Professional Forecasters* and authors' calculations.

The Effect of Monetary Policy on Survey Forecasts of Output

Output growth over coming year
Survey forecast minus benchmark forecast (%)
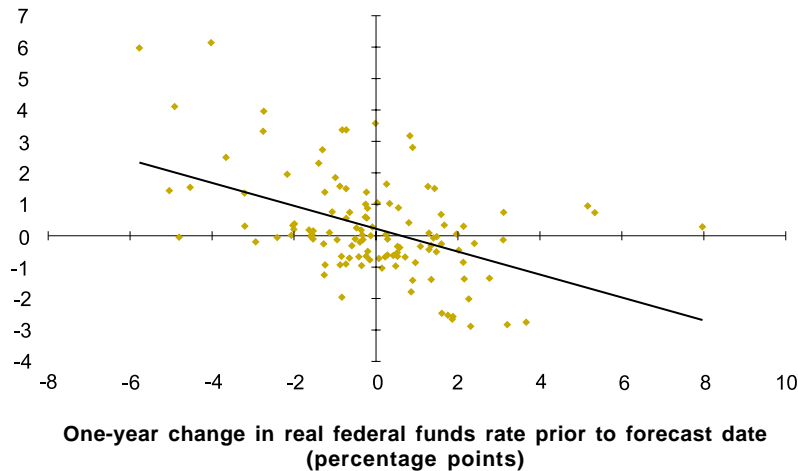
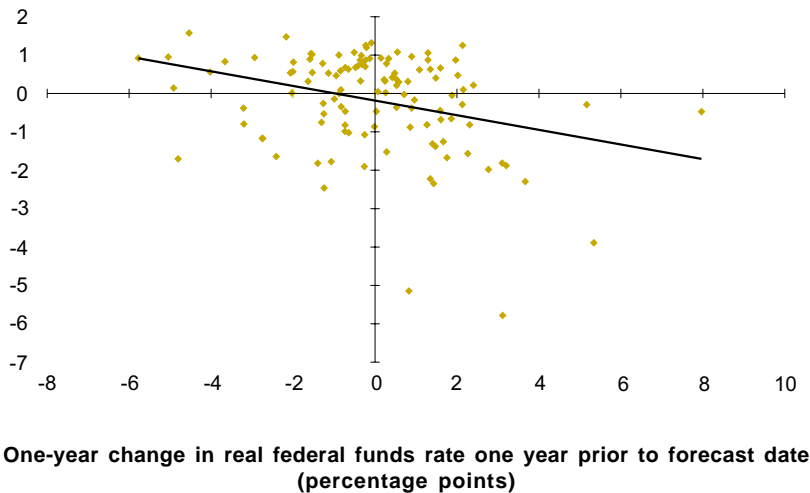**One-year change in real federal funds rate prior to forecast date
(percentage points)**

Note: A linear regression line is plotted.
Source: *Survey of Professional Forecasters* and authors' calculations.

The Effect of Monetary Policy on Survey Forecasts of Inflation

Inflation over coming year
Survey forecast minus benchmark forecast (%)

**One-year change in real federal funds rate one year prior to forecast date
(percentage points)**

Note: A linear regression line is plotted.
Source: *Survey of Professional Forecasters* and authors' calculations.

suggests that economists may incorporate changes in monetary policy into their forecasts while the simple benchmark forecasts can't do so. The same is true of inflation forecasts. But, again, we need to look at changes in monetary policy from a year earlier to see an effect, and again the relationship isn't as clear as it was for output (Figure 4b). This time, however, formal statistical tests show that the negative relationship isn't strong enough to be statistically significant. Thus, monetary policy doesn't significantly affect survey inflation forecasts relative to our benchmark forecasts.

Overall, tighter monetary policy may lead survey forecasts of output growth to be lower than benchmark forecasts, but it doesn't have a statistically significant effect on survey forecasts of inflation relative to benchmark forecasts.

Are the Survey Forecasts Rational? We can also compare the survey forecasts with actual output growth and inflation. This comparison indicates whether the survey forecasts are rational. If they are rational, the survey forecasts should change in response to shifts in monetary policy in the same way that actual output growth or inflation changes. Otherwise, the survey forecasts are irrational — that is, survey respondents could make better forecasts using the information available about monetary policy.

To investigate the rationality of the forecasts, once again we'll look at the forecast errors — the difference between actual output growth or inflation and the survey forecast for those variables. If monetary policy gets tighter (an increase in the real federal funds rate), both actual output growth and the survey forecast for it should decline by the same amount; therefore, the forecast error shouldn't be correlated with monetary policy. The same should be true of easier monetary policy: there should be no relationship between a

measure of monetary policy and the forecast error for output growth. Similar results should hold for inflation.

For output growth, we will look at the forecast error to see if it's correlated with our measure of monetary policy. A scatterplot shows a negative relationship between the forecast error and the measure of monetary policy (Figure 5a), which is statistically significant. The relationship isn't as strong or as large in magnitude as the relationship shown in Figure 3a, which suggests that the survey forecasts do respond to changes in monetary policy, but not enough. In other words, when monetary policy tightens, survey forecasters reduce their forecasts of output growth, but not by enough to match what actually happens. Similarly, easier monetary policy leads forecasters to raise their forecasts of output growth, but not by enough to match reality.

What about inflation forecasts? When we plot the inflation forecast error against past changes in the real federal funds rate, there's a slightly negative relationship (Figure 5b), but it isn't statistically significant. So it appears that forecasters are able to change their forecasts of inflation in response to changes in monetary policy in a rational way.

In summary, survey forecasts of output don't fall enough when monetary policy tightens, but survey forecasts of inflation decline by the right amount. Thus, forecasters are inefficient in forecasting output when monetary policy changes.

CONCLUSIONS

What implications do the results discussed in this article have for how we think about forecasts and monetary policy? If the survey forecasts fail to capture the impact of monetary policy on output growth, then monetary policy could have an additional, indirect effect on the economy; our working paper presents a formal model in which

FIGURE 5a

Are Survey Forecasts of Output Rational?

Output growth over coming year
Actual minus survey forecast (%)
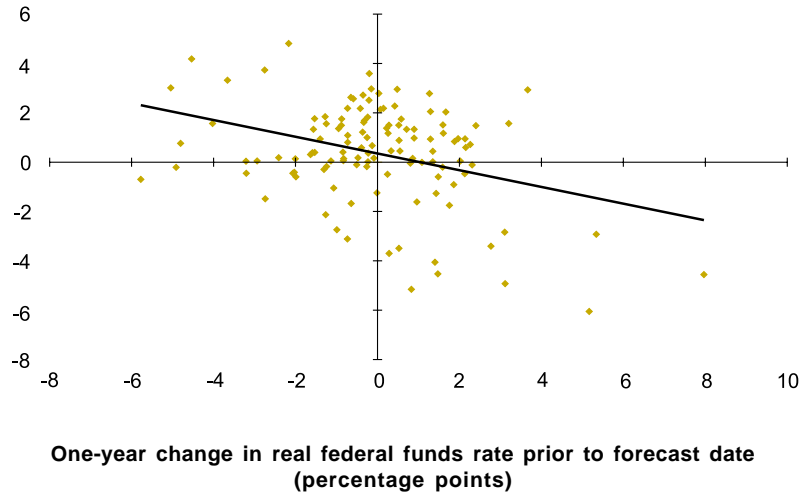
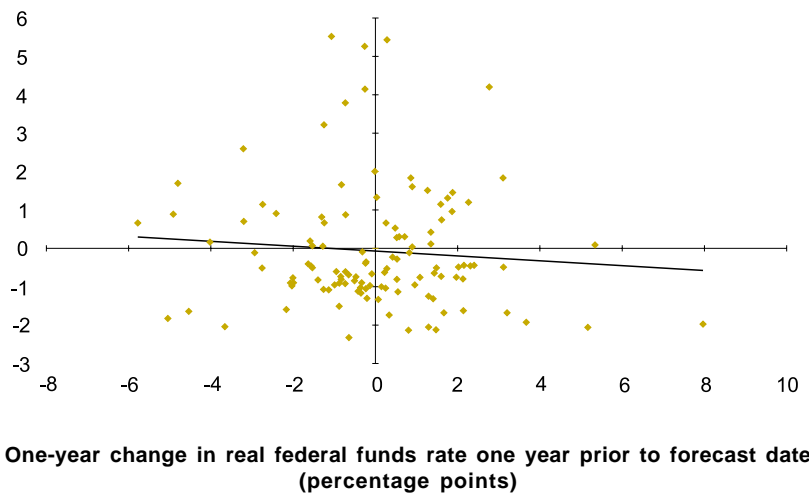**One-year change in real federal funds rate prior to forecast date (percentage points)**

Note: A linear regression line is plotted.
Source: *Survey of Professional Forecasters* and authors' calculations.

FIGURE 5b

Are Survey Forecasts of Inflation Rational?

Inflation over coming year
Survey forecast minus benchmark forecast (%)

One-year change in real federal funds rate one year prior to forecast date (percentage points)

Note: A linear regression line is plotted.
Source: *Survey of Professional Forecasters* and authors' calculations.

this occurs. In particular, some models of the economy assume that a change in monetary policy affects the economy only if the change is a surprise. But even if a change in monetary policy isn't a surprise, its effects may be. Indeed, our evidence suggests that this is so. Even though monetary policy, as measured by a change in the real federal funds rate, is readily observable, forecasts of output don't fully react to it. And this under-reaction provides one possible channel through which monetary policy may affect the economy.

When we examine simple benchmark forecasts, survey forecasts, and actual movements of output growth and inflation, we find three key results. First, the survey forecasts and actual movements of output growth and inflation change when monetary policy changes. Both output growth and survey forecasts of output growth decline when monetary policy tightens and increase when monetary policy eases. Second, there's evidence that forecasts of inflation from the *Survey of Professional Forecasters* are rational; that is, they change as much as they should when monetary policy changes. Third, we've found some evidence that forecasts of output growth from the *Survey of Professional Forecasters* aren't rational, since they don't change as much as they should when monetary policy changes.

This last result is a bit surprising. After all, survey participants provide the best forecasts publicly available for the U.S. economy. Perhaps there have been significant changes in the relationship between output growth and monetary policy, and forecasters will eventually modify their forecasts to reflect that change. But for now, it remains a mystery as to why we find that forecasts aren't fully rational. ⬛

REFERENCES

Ball, Laurence, and Dean Croushore. "Expectations and the Effects of Monetary Policy," Federal Reserve Bank of Philadelphia Working Paper No. 01-12, August 2001.

Croushore, Dean. "Introducing: The Survey of Professional Forecasters," Federal Reserve Bank of Philadelphia *Business Review*, November/December 1993.

Croushore, Dean. "Inflation Forecasts: How Good Are They?" Federal Reserve Bank of Philadelphia *Business Review*, May/June 1996.

Croushore, Dean. "Evaluating Inflation Forecasts," Federal Reserve Bank of Philadelphia Working Paper 98-14, June 1998.

Knowledge Spillovers:
Cities' Role in the New Economy

BY GERALD A. CARLINO

In the smokestack era, cities were centers of manufacturing. What role do cities play in the "new economy"? In this article, Jerry Carlino discusses the link between economic growth and the concentration of people and firms in cities. In particular, he focuses on "knowledge spillovers." These spillovers facilitate the exchange of ideas, promoting creativity and innovation.

Most countries make sustained economic growth a principal policy objective. Although many factors contribute to economic growth, recent research has found that innovation and invention play an important role. Innovation depends on the exchange of ideas among individuals, which economists call knowledge spillovers. For example, a given company's innovation may stimulate a flood of related inventions and technical improvements by other companies.

Recently, some economists have suggested an important link between national economic growth and the concentration of people and firms in cities. The high concentration of people and firms in cities creates an environment in which ideas move quickly from person to person and from firm to firm. That is, dense locations, such as cities, encourage knowledge spillovers, thus facilitating the exchange of ideas that underlies the creation of new goods and new ways of producing existing goods.

Cities and their dense inner-ring suburbs play an important role in the "new economy." In the not-too-distant past, the national economy was based on the production of goods. At the time, cities were good locations for firms because the production of goods was more efficient inside cities than outside them.[1] But manufacturing activity has continually shifted from dense to less dense parts of the country. Consequently, today, our densest cities are important not as centers of manufacturing but as centers of innovation. As economist Janice Madden has pointed out: "To the extent that there is a 'new economy,' it can be described as one in which creativity has become more important than the production of goods." Economist Leonard Nakamura has demonstrated that during the past century, increasingly more workers were "employed in creative activities such as designing, inventing, and marketing new products, and more and more economic activity [was] devoted to creating technical progress."[2] Data from the U. S. Patent Office show that annual applications for patents increased dramatically between the mid-1980s and the mid-1990s. In fact, as we'll see later, most of the patents granted in the 1990s originated in metropolitan areas.

As far back as 1890, Sir Alfred Marshall described cities as "having ideas in the air." In earlier times, cities and their environs contributed to economic efficiency when the economy was based on the production of goods. Today's cities, despite well-publicized drawbacks such as congestion, contribute to the efficient production of knowledge in the new economy.

Jerry Carlino is an economic advisor in the Research Department of the Philadelphia Fed.

---

[1] See my 1987 article for details on how cities increase productivity for manufacturing firms.

---

[2] Leonard Nakamura (2000) shows that in 1900, slightly more than eight of 10 workers produced goods and services. By 1999, the share had steadily declined to slightly more than four of 10.

TWO TYPES OF KNOWLEDGE SPILLOVERS

Economists have identified two types of knowledge spillovers thought to be important for innovation and growth: MAR spillovers and Jacobs spillovers.

MAR Spillovers.[3] In 1890 Alfred Marshall developed a theory of knowledge spillovers that was later extended by Kenneth Arrow and Paul Romer — hence, the name MAR spillovers. According to this view, the concentration of firms in the *same industry* in a city helps knowledge travel among firms and facilitates innovation and growth. Employees from different firms in an industry exchange ideas about new products and new ways to produce goods: the denser the concentration of employees in a common industry in a given location, the greater the opportunity to exchange ideas that lead to key innovations.

Often, the latest information about technological and commercial developments is valuable to firms in the same industry, but only for a short time. Thus, it behooves firms to set up shop as close as possible to the sources of information. For example, many semiconductor firms have located their research and development (R&D) facilities in Silicon Valley because the area provides a nurturing environment in which semiconductor firms can develop new products and new production technologies.

Sometimes, information about current developments is shared informally, as has happened in the semiconductor industry. In her 1994 book, AnnaLee Saxenian describes how gathering places, such as the Wagon Wheel Bar located only a block from

Intel, Raytheon, and Fairchild Semiconductor, "served as informal recruiting centers as well as listening posts; job information flowed freely along with shop talk." Other examples of "high-tech hot spots" include the Route 128 corridor in Massachusetts, the Research Triangle in North Carolina, and suburban Philadelphia's biotechnology research and medical technology industries.

Examples of knowledge spillovers are not limited to the high-tech industry or to the United States. The geographic concentration of the motion picture industry in Los Angeles offers a network of specialists (directors, producers, scriptwriters, and set designers), each of whom focuses on a narrow aspect of movie-making. This network allows easier collaboration, experimentation, and shared learning among individuals and firms. In the medical field, research facilities and teaching institutions have concentrated along York Avenue on Manhattan's Upper East Side, home to Memorial Sloan-Kettering Cancer Center, Rockefeller University and Hospital, and New York Presbyterian Medical Center. Again, this proximity enhances knowledge spillovers among researchers at these institutions.

There are examples of knowledge spillovers in other countries, as well. Economist Michael Porter has cited the Italian ceramics and ski boot industries and the German printing industry, among others, as examples of geographically concentrated industries that grew rapidly through the continual introduction of new technologies.

Jacobs Spillovers. In 1969, Jane Jacobs developed another theory of knowledge spillovers. Jacobs believes

that knowledge spillovers are related to the *diversity* of industries in an area, in contrast to MAR spillovers, which focus on firms in a common industry. Jacobs argues that an industrially diverse urban environment encourages innovation because it encompasses people with varied backgrounds and interests, thereby facilitating the exchange of ideas among individuals with different perspectives. This exchange can lead to the development of new ideas, products, and processes.

As John McDonald points out, both Jane Jacobs and John Jackson have noted that Detroit's shipbuilding industry was the critical antecedent leading to the development of the auto industry in Detroit. In the 1820s Detroit mainly exported flour. Because the industry was located north of Lake Erie along the Detroit River, small shipyards developed to build ships for the flour trade. This shipbuilding industry refined and adapted the internal-combustion gasoline engine to power boats on Michigan's rivers and lakes.

As it turned out, the gasoline engine, rather than the steam engine, was best suited for powering the automobile. Several of Detroit's pioneers in the automobile industry had their roots in the boat engine industry. For example, Olds produced boat engines, and Dodge repaired them. In addition, a number of other industries in Michigan supported the development of the auto industry, such as the steel and machine tool industries. These firms could produce many of the components required to produce autos.

LOCAL COMPETITION

In addition to spillovers, economists have debated the effects of

---

[3] Edward Glaeser, Hedi Kallal, Jose Scheinkman, and Andrei Shleifer, who coined the term MAR spillovers, pulled together these various views on knowledge spillovers in their 1992 article.

> Examples of knowledge spillovers are not limited to the high-tech industry or to the United States.

competition on the rate of innovation and growth: some say more competitive markets innovate faster, and others argue that monopoly encourages innovation. In a classic article in 1961, Benjamin Chinitz contrasted Pittsburgh, which, at the time, was heavily specialized in a few industries and dominated by large plants and firms, with New York City's more diverse and competitive industrial structure. Chinitz suggested that because cities such as Pittsburgh have fewer entrepreneurs per capita, they produce fewer innovations than cities such as New York.

Similarly, Jacobs also believes that the rate of innovation is greater in cities with competitive market structures. According to her, local monopolies stifle innovation whereas competitive local environments foster the introduction of new methods and new products.

In addition, Michael Porter has stated that when local economies are competitive, the innovations of local firms are rapidly adopted and improved by neighboring firms. In contrast, local monopolists tend to rest on their laurels rather than risk innovation.

Alternatively, according to Glaeser and co-authors, the MAR view predicts that local monopoly is superior to local competition because innovating firms recognize that neighboring firms may imitate their ideas without compensation. Therefore, firms in locally competitive environments may invest less in research and development because they do not reap the full benefit of such investment. Thus, local monopoly may foster innovation because firms in such environments have fewer neighbors that will imitate them.

WHAT'S THE EVIDENCE?

In 1991 Paul Krugman noted the difficulty of measuring knowledge spillovers: "Knowledge flows are invisible; they leave no paper trail by which they may be measured and tracked." In a 1993 study, however, Adam Jaffe, Manuel Trajtenberg, and Rebecca Henderson pointed out that "knowledge flows do sometimes leave a paper trail" in the form of patented inventions. Thus, studies of the importance of knowledge spillovers on local inventiveness have relied on patent data. While data on patents imperfectly reflect innovation, they may be the best available measure of inventiveness.[4] For an invention to be patented, it must be useful and novel, and it must represent a significant extension of existing products.[5]

Observing the location of patent originations leads to an important finding: patenting is largely a metropolitan phenomenon. During the 1990s, 92 percent of all patents were granted to residents of metropolitan areas, although only about three-quarters of the U.S. population resides in metropolitan areas.[6] San Jose, California, ranked first both in the number of total patents awarded and in patents per capita. During the 1990s, the San Jose metropolitan statistical area (MSA) averaged almost 18 patents for every 10,000 people, compared with 2.5 patents for every 10,000 people nationally (see Table).[7] The Philadelphia MSA ranked seventh in total patents awarded during the past decade, but only 71st in the number of patents per capita (only three patents per 10,000 people — but that's still 20 percent higher than the national average).[8]

Historical data also show that patent originations are concentrated in cities. In 1966, Allen Pred examined U.S patent data for the mid-19th century and found that patent activity in the 35 principal cities at that time was four times greater than the national average. In 1971 Robert Higgs found that the number of patents issued in the U.S. during the period 1870-1920 was positively related to the level of urbanization.

Among the information contained in a patent are references or citations to previous relevant patents. An examiner at the U.S. Patent and Trademark Office determines which citations a patent must include. For example, if a new patent cites a previous one, that indicates that the older patent contains information on which the newer patent has built. Jaffe, Trajtenberg, and Henderson looked at the propensity of new patents to cite patents that had originated from the same location. They found that a new patent is five to 10 times more likely to cite patents from the same metropolitan area than one would expect, even after eliminating those that are from the same firm. They also found that location-specific information spreads out slowly, making geographic access to that

---

[4] Some innovations are not patented, and patents differ enormously in their economic impact. Nonetheless, patents are a useful measure of the generation of ideas in cities.

[5] See Robert Hunt's article for a succinct review of the patenting process.

[6] The geographic distribution of patents is based on the residence of the inventor whose name appears first on the patent and not the location of the inventor's employer.

[7] MSAs are statistical constructs used to represent integrated labor-market areas that consist of counties containing a central city of at least 50,000 people along with any contiguous counties if such counties meet certain economic criteria. See the article by Bronwyn Hall and Rosemary Ziedonis for an examination of the patenting behavior of 95 U.S. semiconductor firms during the period 1979-95.

---

[8] Maryann Feldman and David Audretsch used the U.S. Small Business Administration's innovation database, which consists of innovations compiled from new product announcements in manufacturing trade journals. They found that in 1982 only 150 of the innovations (4 percent) covered by their data set occurred outside of metropolitan areas. Almost one-half of all innovations occurred in four metropolitan areas: New York (18.5 percent), San Francisco (12 percent), Boston (8.7 percent), and Los Angeles (8.4 percent).

knowledge important to firms. They took these findings as evidence of knowledge spillovers in metropolitan areas.

Estimating the Effect of Urban Density on Patenting. While economists believe that denser areas promote knowledge spillovers that foster innovations, past studies have not looked at the relationship between density and innovation. To investigate this relationship, we need a measure of local employment density. Employment density varies enormously within an MSA. Typically, employment density is highest in the central business district (CBD) of an MSA's central city and generally falls off as we move away from the CBD. An urbanized area is defined as the highly dense area within an MSA.[9] If knowledge spillovers are important, it's likely that urbanized areas with high-employment density would account for most of them.

So ideally, we want to use employment density in the urbanized area of the MSA to investigate the relationship between density and innovation. While we can measure the size of the urbanized part of an MSA, employment data are not available for urbanized areas of MSAs. So we used two alternative measures for local employment density. Our first measure

---

[9] The Census Bureau defines an urbanized area as one with a total population of at least 50,000, consisting of at least one large central city and a surrounding area with a population density greater than 1000 people per square mile.

for local employment density assumes that all employment in an MSA is located within the MSA's urbanized area. This assumption means that our first measure overstates both employment and local employment density. Our second measure is the ratio of employment in the county containing the MSA's central city to square miles in the urbanized area of the MSA. Since the urbanized area is defined to include the MSA's central city *and* the highly dense surrounding areas, our second measure understates both employment and employment density in urbanized areas. By using these alternative measures for local employment density, we believe that the two estimates of the effect of local employment density on the rate of patenting obtained in our

TABLE

Top 50 MSAs' Per Capita Patent Activity in the 1990s

MSA Name	Patents per 10,000 People In the MSA	MSA Name	Patents per 10,000 People In the MSA	MSA Name	Patents per 10,000 People In the MSA
San Jose, CA	17.6	Minneapolis, MN	5.7	Santa Barbara, CA	4.2
Boise City, ID	14.1	Santa Cruz, CA	5.6	Hamilton, OH	4.2
Rochester, NY	13.0	Albany, NY	5.5	San Diego, CA	4.1
Boulder, CO	11.2	Raleigh, NC	5.5	New Haven, CT	4.1
Trenton, NJ	10.5	Brazoria, TX	5.2	Portsmouth, NH	4.1
Burlington, VT	9.0	Manchester, NH	5.2	Lafayette, IN	4.0
Rochester, MN	9.0	Boston, MA	5.1	Rockford, IL	4.0
Poughkeepsie, NY	8.8	Binghamton, NY	5.1	Cincinnati, OH	3.9
Ann Arbor, MI	8.3	Newark, NJ	4.5	Hartford, CT	3.8
Austin, TX	8.0	Kokomo, IN	4.5	Monmouth-Ocean, NJ	3.8
Middlesex, NJ	7.7	Madison, WI	4.5	Akron, OH	3.8
Wilmington, DE	7.5	New London, CT	4.4	Allentown, PA	3.8
Lake County, IL	7.1	Oshkosh, WI	4.4	Greeley, CO	3.8
Saginaw Bay, MI	7.0	Anaheim, CA	4.4	Seattle, WA	3.8
Ft. Collins, CO	7.0	Cedar Rapids, IA	4.3	Kalamazoo, MI	3.8
Bridgeport, CT	6.7	Elmira, NY	4.3	Sheboygan, WI	3.8
San Francisco, CA	5.8	Oakland, CA	4.2		

analysis will capture the true effect of density on innovation.[10]

Data for the 1990s on 270 MSAs reveal a positive association between patents per capita and local employment density.[11] But, as we just discussed, other characteristics of the local economy (such as its industrial structure and its competitiveness) can also affect the number of patents. A standard statistical technique, called multiple regression analysis, can be used to identify the factors that best explain MSA differences in patents per capita. We considered the effects of a wide range of factors — such as the number of employed people in the MSA (or MSA employment), R&D spending in science and engineering programs at colleges and universities (university R&D), the share of large firms (1000 or more employees), and educational attainment of the population — on patents per capita in metropolitan areas to determine how the number of patents per capita during the 1990s was affected by metropolitan employment density in 1989 (see the Appendix).

Density. During the 1990s, patenting was significantly greater in MSAs with denser local economies. For example, the number of patents per capita was, on average, 20 percent to 30 percent higher in an MSA whose local economy was twice as dense as that of another MSA. Since local employment density varies by more than 2000 percent across locations in the sample, the implied gains in patents per capita due to urban density are substantial. For example, in 1989, the average urbanized area in our sample had about 1500 jobs per square mile (assuming all jobs in the MSA are located inside its urbanized area). Toledo, Ohio; Eugene, Oregon; and Omaha, Nebraska are three MSAs with local employment density at about this average level. These three MSAs averaged 1.8 patents per 10,000 people during the 1990s. If their local employment density were to double, the statistical model predicts that patents would rise, on average, to 2.3 per 10,000 people. Thus, these findings are consistent with the widely held view that the nation's densest locations — its central cities and their dense inner-ring suburbs — play an important role in creating the flow of ideas that generate innovation and growth.

However, before we can reach a definitive conclusion, we must remember that the rate of patenting may be greater in denser locations for reasons other than knowledge spillovers. For example, it's possible that in urban areas it's harder to keep information secret, so firms resort to patents. Wesley Cohen, Richard Nelson, and John Walsh examined this possibility in a study, referred to as the Carnegie Mellon Survey (CMS), which was based on a 1994 survey of R&D at 1478 manufacturing firms. Results of the CMS show that manufacturing firms typically protect the profits from their innovations with a variety of mechanisms, including patents, secrecy, and first-to-market advantages. Furthermore, the majority of manufacturing firms surveyed indicated that they rely on secrecy and first-to-market advantages more heavily than patents.

More important for our purposes, surveyed firms indicated that concern over information disclosed in patents is a major reason many choose not to pursue a patent. Current laws require patents to describe an invention in precise terms. In addition, there are high fixed costs associated with preparing a patent application (such as legal fees and the cost associated with patent searches). Secrecy, however, avoids these fixed costs, but preventing disclosure of secret information incurs expenses. Although the CMS does not consider the location of the firms in its sample, its findings nonetheless suggest that firms may be forced to rely on patenting to a greater extent in dense areas because it is harder and more costly to maintain secrecy there than in less dense areas. Thus, it may be this increased difficulty in maintaining secrecy, and not knowledge spillovers, that accounts for the positive correlation between patents per capita and metropolitan density.

Unfortunately, we cannot distinguish between the effects of knowledge spillovers and those of secrecy in our empirical model.[12] While the inability to maintain secrecy in dense locations may account for some portion of the positive association between patents per capita and density, it is unlikely that it would completely "crowd out" the effects of knowledge spillovers.

> Before we can reach a definitive conclusion, we must remember that the rate of patenting may be greater in denser locations for reasons other than knowledge spillovers.

---

[10] See the Appendix for details on how the local employment density variables are constructed.

[11] The simple correlation between the logarithm of patents per capita and the logarithm of local employment density is moderately positive (0.50) and statistically significant.

[12] At this time, data that would allow us to discern the role of knowledge spillovers and that of secrecy in patent activity in dense local areas are not publicly available.

and firms in cities fosters innovation and, along with the findings of other studies, offer little support for the MAR view that specialization and local monopoly foster innovation. The evidence is mixed on Jacobs' view: While we find little evidence that the rate of innovation is greater in diverse and locally competitive environments, studies by Glaeser and co-authors and by Feldman and Audretsch, however, report results favorable to this view.

CONCLUSION

The extraordinary recent growth in productivity and jobs in the United States has been attributed in part to innovation. The empirical work we discuss in this article has shown that patent activity is positively related to the density of an MSA's highly urbanized area (the portion containing the central city). Our findings suggest that dense urban areas, such as central cities, foster knowledge spillovers, which are important in the generation of new ideas that lead to new products and new ways to produce existing products.

Given the role that dense geographic locations may play in promoting innovation, the postwar decline of the nation's dense central cities relative to their less dense suburbs should be a concern to both local and national policymakers. In fact, in a 1997 study, Joe Gyourko and Dick Voith showed that many central cities have experienced not only declines in economic activity relative to their suburbs but absolute declines as well. Sound urban policies are necessary to make the most of the growth potential that the central cities of the nation's metropolitan areas offer. But local and national policies have often contributed to the suburbanization of jobs and lowered the employment density of central cities. In doing so, they may have weakened the economy's ability to innovate and may ultimately lead to slower growth. ⊕

Industrial Specialization. Even if we accept the view that dense local areas serve as centers for the exchange of ideas, we come back to the issue of whether the rate of exchange is enhanced in industrial environments that are diverse (for example, New York City) or in more specialized ones (for example, Silicon Valley). Feldman and Audretsch's 1999 study, which used the U.S. Small Business Administration's innovation database, focused on innovative activity for particular industries within specific MSAs. They found less industry-specific innovation in MSAs that specialized in a given industry, a finding that supports Jane Jacobs' diversity thesis. Glaeser and co-authors provided indirect evidence by looking at employment growth between 1956 and 1987 across specific industries in a given city. They found that industrially diversified areas grew more rapidly than specialized areas.

Conversely, in our empirical work, we found little evidence that diversity, or the lack of it, was an important factor in determining the rate of patenting activity in metropolitan areas in the 1990s.

Competition. Finally, we look at the evidence on whether the creation of ideas is greater in competitive local environments characterized by many small firms than in local economies dominated by a few large firms. Feldman and Audretsch found that local competition is more conducive to innovative activity than is local monopoly. More indirect evidence on this issue is offered by Glaeser and co-authors' finding that local competition is more conducive to city growth than is local monopoly. Counter to these studies, and to the views of Chinitz and Jacobs discussed earlier, our empirical findings show that, overall, patenting is not related to local competition or the lack of it.[13]

In sum, our findings suggest that the high concentration of people

---

[13] In our empirical model, we examine the rate of local patenting and a number of other characteristics of the local economy (such as the level of employment in an MSA, the relative importance of large firms in an MSA, the percent of total MSA employment in manufacturing, and the percent of an MSA's population with a college education). The level of MSA employment, the relative importance of large firms in an MSA, the percent manufacturing in an MSA, and the percent college educated in an MSA were associated with significantly higher rates of MSA patenting during the 1990s (see the Appendix for details).

REFERENCES

Carlino, Gerald A. "Productivity in Cities: Does Size Matter?" Federal Reserve Bank of Philadelphia *Business Review* (November/December 1987).

Carlino, Gerald A., Satyajit Chatterjee, and Robert Hunt. "Knowledge Spillovers and the New Economy of Cities," Working Paper 01-14, Federal Reserve Bank of Philadelphia (2001).

Chinitz, Benjamin. "Contrasts in Agglomeration: New York and Pittsburgh," *Papers and Proceedings of the American Economic Association*, 51 (May 1961), pp. 279-89.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh. "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (OR NOT)," Working Paper 7552, National Bureau of Economic Research (2000).

Feldman, Maryann P., and David B. Audretsch. "Innovation in Cities: Science-Based Diversity, Specialization and Localized Competition," *European Economic Review* (1999) pp. 409-29.

Glaeser, Edward, Hedi Kallal, Jose Scheinkman, and Andrei Shleifer. "Growth in Cities," *Journal of Political Economy*, 100 (1992), pp. 1126-53.

Gyourko, Joseph, and Richard Voith. "Does the U.S. Tax Treatment of Housing Promote Suburbanization and Central City Decline?" Working Paper 97-13, Federal Reserve Bank of Philadelphia (1997).

Hall, Bronwyn H., and Rosemarie H. Ziedonis. "The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-95," *RAND Journal of Economics*, 32 (2001), pp. 101-28.

Higgs, Robert. "American Inventiveness, 1870-1920," *Journal of Political Economy*, 79 (1971) pp. 661-67.

Hunt, Robert M. "You Can Patent That? Are Patents on Computer Programs and Business Methods Good for the New Economy?" Federal Reserve Bank of Philadelphia *Business Review* (First Quarter 2001).

Jackson, John. "Michigan" in R. Scott Fosler, ed., *The New Role of American States*. New York: Oxford University Press (1988).

Jacobs, Jane. *The Economy of Cities*. New York: Vintage Books (1961).

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 108 (1993) pp. 577-98.

Kortum, Samuel, and Josh Lerner. "Stronger Protection or Technological Revolution: What Is Behind the Recent Surge in Patenting?" Working Paper 6204, National Bureau of Economic Research (1997).

Krugman, Paul. *Geography and Trade*. Cambridge: MIT Press (1991).

Madden, Janice. "Creating Jobs, Keeping Jobs, and Losing Jobs: Cities and the Suburbs in the Global Economy," unpublished manuscript, 2000.

Nakamura, Leonard I. "Economics and the New Economy: The Invisible Hand Meets Creative Destruction," Federal Reserve Bank of Philadelphia *Business Review* (July/August 2000).

McDonald, John F. *Fundamentals of Urban Economics*. Upper Saddle River, NJ: Prentice Hall, 1997.

Porter, Michael E. *The Competitive Advantage of Nations*. New York: Free Press (1990).

Pred, Allen R. *The Spatial Dynamics of U.S. Urban Industrial Growth, 1800-1914*. Cambridge: Harvard University Press (1966).

Saxenian, AnnaLee. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge: Harvard University Press (1994).

The variables that were considered in the empirical model are those thought to affect patenting at the MSA level, as discussed in the text.[a]

$$\ln (\text{Patents per Capita}_i) = C + a_1 \ln (\text{MSA Employment}_i) + a_2 \ln (\text{Employment Density}_i) + a_3 \ln (\text{University R\&D}_i)$$
$$+ a_4 \text{Large Firms}_i + a_5 \text{Manufacturing Share}_i + a_6 \text{College Educated}_i + a_7 \ln (\text{Industrial Specialization}_i) + a_8 \ln (\text{Competition}_i)$$
$$+ a_9 \text{Employment Growth}_i$$

where

$\text{Patents per Capita}_i$ = Patents per capita, annual average for the period 1990-99 in MSA i

MSA Employment_i = 1989 level of private nonfarm employment in MSA i

$\text{Employment Density}_i$ = The density of employment in 1989 in the ith MSA's urbanized area

Two alternative measures are used: in model (1) employment density = MSA employment divided by square miles in the MSA's urbanized area; in model (2) employment density = employment in the county containing the MSA's central city divided by square miles in the urbanized area.

$\text{University R\&D}_i$ = University R&D spending in science and engineering programs, annual average for the period 1989-91 in MSA i

Large Firms_i = Percent of firms with 1000 or more employees in 1989 in MSA i

$\text{Manufacturing Share}_i$ = Manufacturing share of total employment in MSA i, in 1989

$\text{Percent College Educated}_i$ = Percent of 1990 population with at least a college degree in MSA i

$\text{Industrial Specialization}_i$ = the Herfindahl index = $\sum_{j=1}^{7} (s_{j,i})^2$, where $s_{j,i}$ is the share of employment in industry j in MSA i

Competition_i = Total number of firms in MSA i divided by total employment in MSA i

$\text{Employment Growth Rate}_i$ = employment growth rate in MSA i during the period 1979-89.

The dependent variable refers to patents per person in the MSA averaged over the period 1990-99, whereas the independent variables are at 1989 or roughly beginning-of-the-period values. This reduces the simultaneity and reduces concerns about direction-of-causation issues, since the value of the dependent variable that is averaged over the 1990s is not likely to affect beginning-of-period values of the independent ones. Employment size in 1989 is included because other researchers have found that innovative activity increases with MSA size.

Research and development (R&D) spending in science and engineering programs at colleges and universities is included separately, since many authors have found spillovers from such spending and innovative activity at the local level. Similarly, since large firms tend to spend proportionately more on private R&D than do smaller firms, the percentage of an MSA's firms with 1000 or more employees is included separately to capture the presence of large firms on patent activity. The percent of an MSA's population with at least a college degree is included to separately account for the role of educational attainment in patenting.

The share of MSA employment accounted for by each of seven industries is used to calculate the Herfindahl index of industry specialization.[b] Squaring each industry's share of employment, s_i,

means that larger industries contribute more than proportionately to the overall value of the index. Thus, as the index increases in value for a given MSA, this implies that the MSA is more highly specialized or less diversified industrially. Following Glaeser et al. (1992), we use the total number of firms per worker in an MSA as a measure of competition; that is, an MSA is taken as locally competitive if it has many firms per worker. Finally, employment growth during the period 1979-89 is included to control for any independent effect that local growth may have had on patent activity.[c]

[a] For additional details, see the working paper by Carlino, Chatterjee, and Hunt.

[b] The seven industries are manufacturing; transportation, communications, and public utilities; wholesale trade; retail trade; services; finance, insurance, and real estate; and other industries. Construction's share of private nonfarm employment was not included in the calculation of the index because of disclosure problems associated with this variable for some MSAs in our sample.

[c] We included dummy variables in both versions designed to see if specific regions of the country contributed more or less to MSA patenting. Each MSA was classified into one of eight broad regions (New England, Mideast, Great Lakes, Plains, Southeast, Southwest, Rocky Mountain, and Far West). We found that MSA patents were higher in the Mideast and Great Lakes regions relative to the Southeast region; the coefficients for the other regions were not statistically significant.

APPENDIX

The model was estimated using ordinary least squares methods with White robust standard errors to take heteroskedasticity into account.

As indicated in the text, one problem is that employment data for urbanized areas are not available. Therefore, we must estimate it. In model (1) we assume that all employment in an MSA is located within its urbanized area. This assumption overstates both employment and employment density in urbanized areas. In model (2) we assume that all employment in an MSA is located within the county that contains the MSA's central city. This assumption understates both employment and employment density in urbanized areas.[d]

_____

[d] On average, the county containing an MSA's central city accounts for 84 percent of MSA employment.

The results of the regression are presented in the table on the next page. As the results of both models show, the effect of employment density on patenting is positive and highly significant. These findings suggest the importance of close spatial proximity in promoting spillovers and fostering innovation. A number of other variables in the model have the expected positive association with the rate of MSA patenting, including MSA employment size, percent of MSA firms with 1000 or more employees, percent of MSA employment in manufacturing, and the percent of MSA population with a college education. The coefficient on the Herfindahl index is not statistically significant, suggesting that an MSA's degree of industrial specialization does not have a significant impact on MSA patenting. Similarly, the variable firms per employee is not significant, suggesting that competitiveness of the local economy does not appreciably affect MSA patenting activity. One anomaly is that university R&D spending has the wrong sign (negative, which suggests that increased spending by local universities on R&D in science and engineering programs is associated with fewer patents per capita in an MSA), but it is not significant. Finally, the R^2 statistic, measuring the goodness of fit, shows that the models explain a little more than 60 percent of the variation in MSA patents per capita (this is a good fit for a cross-MSA model).

The Determinants of Patents Per Capita[a]

	(1)[b]	(2)[c]
Urbanized area employment density (MSA employment)	0.3058**	
Urbanized area employment density (central city's county)		0.2056**
1989 Employment	0.2985**	0.3368**
University R&D spending	-0.0086	-0.0102
Percent of firms with 1000 or more employees	202.1*	227.9**
Percent mfg.	3.66**	4.12**
Percent college educated	6.63**	6.60**
Herfindahl index	1.4785	1.8249
Firms per employee	0.5298	0.5654
Employment growth, 1979-89	0.1018	0.1253
Constant	-13.8**	-13.1**
No. of Obs.	270	257
R^2	0.6138	0.6169

* and ** indicate statistically significantly different from zero at 5 percent and 1 percent levels, respectively.

[a] Both models include a set of dummy variables to account for the MSA's region.

[b] In model (1) employment density = MSA employment divided by square miles in the MSA's urbanized area.

[c] In model (2) employment density = employment in the county containing the MSA's central city divided by square miles in the MSA's urbanized area.
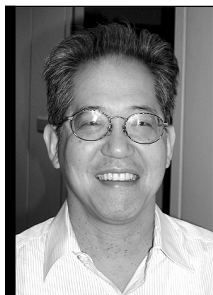
Investing in Intangibles:
Is a Trillion Dollars Missing from GDP?

BY LEONARD NAKAMURA

I n the 1990s, Americans saved less, but they became wealthy at an astonishing rate. What underlies this paradox of a lower savings rate coupled with increased wealth? As Leonard Nakamura states in this article, the short answer is capital gains. Stock-market capital gains are excluded from our measures of national income, yet they account for about half of the increase in American households' net worth in the past two decades. Nakamura discusses the pros and cons of including capital gains in national income accounts.

Writing *David Copperfield* in 1849, Charles Dickens put these rueful words into the mouth of the feckless Mr. Micawber: "Annual income twenty pounds, annual expenditure nineteen, nineteen six, result happiness. Annual income twenty pounds, annual expenditure twenty ought and six, result misery."[1] The inability to save leads to the poorhouse, as Dickens well knew,

since his father's debts had done just that to his family. But in the 1990s Americans saved less and less, according to official U.S. statistics. Yet far from being miserable, they became wealthy at an astonishing rate.

What underlies this paradox of a small saving rate in tandem with increased wealth? The short answer is: capital gains. Specifically, saving and wealth gains diverge because of a convention in the U.S. income accounts that makes a good deal of sense. Because capital gains are so volatile, the national income accounts include only part of investment income: dividends and interest payments.[2] Capital gains

are excluded, yet capital gains from the stock market have been responsible for about half of the increase in the net worth of American households in the past two decades.[3] This rise in capital gains has occurred because firms can reward shareholders either with dividends or with capital gains, and U.S. corporations have been retaining more of their earnings in the form of intangible investment and not paying them out in dividends.[4]

The official measure of U.S. household saving, the personal saving rate, is, like all economic statistics, a compromise between a theoretical ideal and the practical limitations of existing data.[5] Ideally, we expect key statistics, such as the saving rate, real GDP growth, and consumer price inflation, to convey important information as clearly

---

Leonard Nakamura is an economic advisor in the Research Department of the Philadelphia Fed.

---

[1] Micawber's money is in pounds, shillings, and pence. There were 20 shillings to a pound and 12 pence to a shilling.

[2] See the article by Richard Peach and Charles Steindel for an interesting discussion of this problem and the importance of realized capital gains (capital gains that investors have received by selling their investments and, thus, can be used to pay for consumption).

[3] The market value of domestic corporate equities rose $12 trillion, from $2 trillion at the end of 1979 to $14 trillion at the end of 2000, in 1996 dollars. During that time, the total net worth of U.S. households (which hold almost all of domestic equities) rose $23 trillion, from $15 trillion to $38 trillion. By contrast, real estate holdings of U.S. households rose by about $6 trillion during this period.

[4] See the article by Eugene Fama and Kenneth French.

[5] Reports about saving usually focus on household saving, that is, personal saving. Personal saving is defined as disposable (that is, after-tax) personal income less personal outlays (personal consumption expenditures plus transfers abroad). Personal income includes dividends and net interest payments from corporations, but not capital gains. It also includes wages and salaries, employment benefits like health insurance, noncorporate income such as proprietors' income and rental income, and net transfers from government, such as Social Security benefits.

as possible. In this ideal, a very low saving rate should not be compatible with substantial and sustained creation of wealth.

Let's look at Mr. Micawber again. If he has a steady income of 20 pounds a year and no capital assets, determining his income is simple: 20 pounds. And regular income, such as paychecks, are generally what our statistics measure. But what if Micawber owns some stock? Then measuring his income is no longer so simple. If his stock rises in value from 10 pounds to 11 pounds, should Micawber's income be calculated as 20 pounds or 21? And how should Micawber report his income when his paper profits disappear and turn into a paper loss? A key question then for Micawber's budget problem is: given that stock prices go up and down, how much of the gain can he rely on, and thus, how much can he afford to spend?

If we include capital gains in personal saving, the U.S. saving rate, properly measured, has generally risen rather than fallen.[6] But improving our statistical measures is by no means straightforward. Why? Fortunately for our economic well-being, but unfortunately for the credibility of our statistical measures, economic activity is increasingly concerned with the creation of new products. This type of economic activity is difficult to capture accurately in our economic measures. In fact, given how we construct the personal saving rate for the United States, a low or even negative saving rate is likely to coexist with substantially accelerated creation of wealth.

Shedding some light on this paradox of diminished saving and

increased wealth and why it's difficult to eliminate it is the purpose of this article.

RESOLVING THE PARADOX

Why did wealth accelerate? Were we lucky? Or were we actually saving more, but miscounting it? To the extent that saving was undercounted, we should expect wealth gains to be

If we include capital gains in personal saving, the U.S. saving rate, properly measured, has generally risen rather than fallen.

sustainable in the future. But if all the gain was due to good luck, we must reduce our consumption relative to our incomes if we want our wealth to continue to grow over the long run.

What we save can be measured as the resources we, as a society, put toward the future — the labor and capital devoted to new investment rather than immediate consumption. But investing is often risky: an investment sometimes returns a multiple of the original investment, but sometimes much less. When estimating GDP, we can calculate investment by measuring how much we invested or by measuring the outcome of the investment, that is, the net wealth generated.

Recently, in fact, the dot-com bubble gave us an object lesson in the difference between resources invested and wealth created, since much of the investment made in this sector has come to naught. This outcome is, unfortunately, all too typical when we try to create new products. The risk intrinsic to investing in new products means that the outcome of the investment and the

dollars invested are very likely to be different.

Intangible assets are primarily derived from the property rights to which firms become entitled when they create new goods and services. We can use the analogy of cooking to divide economic activities into the creation of new menu items (creating recipes) and the actual production of food ready for the diner (following recipes). Intangible investment is the creation of recipes, and the intangible asset created — the result of the recipe — is the patent, copyright, trade secret, or brand name that protects the creator's right to exclusively reproduce or use the recipe. When a private corporation uses this right to sell new items, it can charge a monopoly price to consumers, and thus — if the new item is highly desirable — earn outsize profits on these assets, profits that repay the cost of creating the item. In turn, once private investors recognize the value of the creation, the corporation's stock-market value will rise, causing its shareholders' wealth to increase.

Even if we include the effects of the recent downturn in the stock market, in the past two decades, the wealth of U.S. households has increased dramatically, and much of this increase has taken the form of these stock-market capital gains due to successful investments in intangible assets.

Taking account of this investment has become more pressing because investment in intangible assets has become a bigger part of the U.S. economy. In the past, most business investment took the form of tangibles: equipment such as trucks, computers, and typewriters; and structures such as office buildings, shopping malls, and homes. But in the past 20 years, accelerating investment in intangibles — investments that result in patented discoveries like Viagra and Celebrex or copyright-protected products such as Windows2000, Pentium, and Harry

---

[6] I argue this case in my working paper, "What Is the U.S. Investment in Intangibles? (At Least) One Trillion Dollars a Year!"

Potter — has increasingly driven U.S. firms and raised their economic value. However, investing in intangibles is much riskier than investing in tangibles. And taking that riskiness into account is not easy.

There are two different approaches we can take: measuring outcomes or measuring more intangible investment. And each approach gives us a different answer.

If we measure outcomes, we ought to include stock-market capital gains as part of income. This gives us a measure that is useful in tracking wealth. However, including these capital gains in our definition of income makes income much more volatile than other measures of economic activity, such as employment.[7] Also, if we include capital gains in income, the personal saving rate, on average, would have been much higher over the 1990s, but also more volatile.

An alternative is to include more intangible investment as measured by the cost of the inputs — the resources used in this investment — rather than counting capital gains, which are a measure of the success of the investment. If we adopted this approach, measured corporate retained earnings and private gross saving would be larger, but the personal saving rate would likely remain low.[8]

---

[7] Some of this volatility reflects fundamental volatility in the economy, while some of it reflects uninformative noise. Disentangling the two sources of volatility is very difficult, particularly over short periods.

[8] Total national gross saving includes personal saving, corporate gross saving, and government saving. Corporate gross saving includes retained earnings and depreciation allowances. As we include more intangibles in gross investment, both measured retained earnings and depreciation allowances will rise. Only when dividends rise (shifting saving from the corporate sector to private households) to fully reflect increases in corporate profits will the personal saving rate return to its longer run average.

CLASSIFYING OUTPUT AND MEASUREMENT

Measuring economic output entails a fundamental issue: how to avoid double-counting it. For example, when a consumer buys two scrambled eggs at a diner, we count the tab as part of output. We don't want to count separately the feed that the hen ate because the cost of feed is part of what the consumer paid for. The feed is an intermediate output used in producing

By its very nature, investment takes resources that might otherwise have been consumed to create a product whose value will only be fully realized over time.

the final output, scrambled eggs.

This same rationale might be used to exclude saving and investment from measures of national income. We *could* treat investments as intermediate goods because ultimately they are also incorporated into final consumption goods. After all, a truck's value to consumption derives from its role in production: hauling goods that are ultimately consumed. Similarly, without a stove, a short order cook can't make scrambled eggs.

Two Good Reasons for Counting Investment as Part of Output. But one reason we may wish to count investment as part of output is that we could have used the resources that went into investment to simply increase consumption today. By its very nature, investment takes resources that might otherwise have been consumed to create a product whose value will only be fully realized over time. If we fail to include investment as part of output, we undercount the potential productivity of our existing resources and omit the opportunity cost of the investment, that is, what else we could have done with our inputs.

A second reason for counting investment as part of output is that it represents a store of value. Investing in a truck or stove is valuable because these items can be used to help us create more consumables in the future. By counting these investments as part of output, we recognize that when investments succeed, our wealth increases. Our wealth, in turn, will enable us to consume more in the future. Not counting additions to wealth ignores the future output that this wealth could produce.

But Intangibles Have Not Been Counted. Historically, in the U.S. national income accounts, only tangible investments in equipment and structures have been included in our measures of investment. Until very recently, investment in intangibles has been ignored. Intangibles have traditionally been treated just as if they are intermediate goods and services that need not be counted because they are subsequently incorporated into final goods and services. But because intangible investment uses resources to create products whose value is not immediately realized, failing to count it understates both our current ability to produce and our assets. When we incompletely count assets whose purpose is to increase future production, we will be surprised by the extra income earned subsequent to the investment, and profits will grow faster than anticipated.

Beginning in 1998, the Bureau of Economic Analysis (BEA) has included software as the first intangible investment in its measure of GDP. Between 1998 and 2000, measured

business investment in software rose from $140 billion to $183 billion, in current dollars.

Other investments in intangible assets, such as research and development (R&D), movie and book production, designs and blueprints, and the advertising associated with the new products produced, could also be included in output. Because these are important sources of wealth creation, it seems likely that the BEA will eventually do so. In the meantime, official statistics in the United States will continue to understate output and saving.

MEASUREMENT PROBLEMS: GREATER BECAUSE INTANGIBLES ARE RISKIER

A substantial difference between tangibles and intangibles is that the production process for tangibles is much less risky than that for intangibles. When a truck or an oven is produced, the outcome — and its value — are highly predictable. Mass production, by its very nature, churns out multiple, identical copies of the same product. If a firm spends $10 million to equip a factory, the value of that equipment is relatively easy to document.

Mass-produced equipment often has a second-hand market in which the value of the used equipment can be determined. Indeed, in some cases, such as cars and trucks, standard estimates of the value of "pre-owned" equipment are published. Moreover, accountants and auditors can verify the existence of the asset. If the equipment loses its value in the second-hand market, and the purpose for which the equipment was bought turns out to be worthless, the accountant is supposed to write off the investment, deducting it as an expense.

When firms invest in intangibles, on the other hand, the product of the investment is unique and often hard to evaluate objectively. In fact, the product often turns out to be worthless. When a firm invests in producing a design, a movie, or a drug, it hopes to end up with something sufficiently original so that it will have, at least for a time, a monopoly of some segment of the market. For the monopoly to have substantial value, the intangible asset must offer something no other product on the market offers.

But efforts to produce what no one has been able to make before often misfire. For example, many drugs that are promising in theory and that work well in the laboratory or on animals turn out to be unsafe or ineffective for human patients in clinical trials. Other drugs turn out to be worth tens of

A substantial difference between tangibles and intangibles is that the production process for tangibles is much less risky than that for intangibles.

billions of dollars. A large pharmaceutical company may have dozens of drugs in its development pipeline. Generally, less than one in 10 will earn back more than its cost, but that one success may well justify all the failures and make a company's overall research program a success.

Frederic Scherer and Dietmar Harhoff's research on patents issued in the United States and Germany showed that the most valuable 10 percent of patents accounted for between 81 percent and 93 percent of the total value of the sets of patents studied.[9] In their sample of 772 German patents, for instance, the top five — less than 1 percent — accounted for 54 percent of the value of the pool. Thus, a dispropor-

tionate part of the value of all projects is included in a few successful projects.[10]

A firm making a $10 million investment in each of 10 new products may wind up with an asset worth nothing nine times out of 10, but the tenth time may produce an asset worth $100 million. Realizing the long odds against success in intangible investment, accountants have opted to write off intangible investments — acting as if they were intermediate products that did not result in wealth creation. And if the samples in Scherer and Harhoff's study are a good guide, writing off the investment will be the right thing to do in most instances. But the right thing to do most of the time is, on average, the wrong thing to do. Why? Because the few investments in intangibles that do succeed may well be worth more than all other investments put together. In the example above, the firm's 10 investments turn out to be — in all — worth $100 million. So if the firm had written off none of its investments, it would have much more accurately represented its total investment than if it had written off nine out of the 10 – or 10 out of 10, as is current practice.

A Successful Investment in Intangibles: An Example from Pharmaceuticals. As an alternative to current practice, what about measuring inputs? Consider a pharmaceutical company that does research to discover a drug that will cure a previously

---

[9] The studies that Scherer and Harhoff survey include corporate patents, university patents, and pharmaceutical patents.

[10] Technically speaking, these sorts of risks are said to have highly skewed probability distributions.

incurable disease. For example, in June 2000, Eli Lilly announced its belief that Xigris, its treatment for septic shock,

process to make a particular drug. The ascorbic acid will be almost completely used in the year it is purchased, and it is

budget, not adjusted for inflation, was $15.1 billion; carried forward to 2000, this investment had a present value of about $40 billion. Because of its unusual success, Xigris alone could justify much of Eli Lilly's R&D investment for the previous two decades.[13]

> Another important difference between tangible and intangible investing is that the firm that makes tangible capital goods is typically different from the firm that will use them.

would pass its final trials and that its application to the Food and Drug Administration would be successful.[11] Test results indicated that Xigris would save perhaps 20,000 lives annually and earn Eli Lilly as much as $1 to $2 billion annually in profits over the next decade.

Eli Lilly's expenditures on Xigris — including the research that went into discovering its use as a treatment for septic shock, the clinical trials to establish the safety and efficacy of the treatment, and efforts to publicize and market the drug to doctors and medical systems around the world — are investments that will bear fruit in the form of substantial profits over an extended period.

However, our national accounts don't include these expenditures as investments. Instead, these expenditures are treated as expenses — as if they were part of the inputs into products Eli Lilly is currently selling. To draw a parallel, consider two other types of expenditures Lilly might make. One is the purchase of equipment for mass producing a drug. This equipment is considered an investment because it will continue to produce output well after the year of its purchase. Another type of expenditure is the purchase of ascorbic acid, which will be used in a chemical

one of the costs that Lilly rightly expenses in making that particular drug. Similarly, by calling research and development an expense, we are in effect saying that when the R&D is finished, Eli Lilly doesn't possess a valuable asset. And that is surely not the case.[12]

On the day Eli Lilly announced the likely success of its drug (no previous septic shock treatment had been successful), its stock-market value went up $16 billion. Will Xigris' profits justify this increase in value? Given the size of the potential market for the drug and the number of lives it could save, analysts who follow Lilly judged that this single product could well be worth $10 billion or more.

However, Lilly did not invest $16 billion to produce Xigris. Indeed, from 1980 to 1999, Lilly's entire R&D

This example demonstrates that from the perspective of reporting to shareholders, as well as for internal corporate operations, there should be a strong presumption against the premature expensing of intangible investments because doing so understates the profitability of current operations. For example, a corporation might capitalize and depreciate intangible assets according to a predetermined schedule, just as it would a tangible investment. Only when it's clear that a whole group of intangible investments has failed would the corporation write them off as an expense.

Furthermore, this example shows that the resources that go into a risky intangible investment rarely equal its product. The economic resources used in producing an intangible asset will rarely even approximately equal the market valuation of the results of the new product development.

By contrast, in a mass-production economy, input almost always equals output. That is, any given

---

[11] Specifically, Eli Lilly released an announcement that the trial would be closed to new patients earlier than planned.

[12] This represents a fundamental problem in accounting for investment in intangible assets, one probably not entirely solvable using standard accounting treatment of investment. Tangible investments are capitalized, then depreciated. That is, when the expense is first incurred, it is charged to the capital account and not deducted from current revenues. Then, over time, as the tangible asset declines in value, the depreciation is subtracted from current revenues, or expensed. By contrast, since accountants don't want to include as investment assets that cannot be concretely evaluated, intangible assets are expensed when incurred, rather than over time. As corporate investment shifts away from tangibles toward intangibles, current profits become understated. See my 1999 *Business Review* article.

[13] In addition to Xigris, Eli Lilly's research has also produced Prozac, an antidepressant, and Zyprexa, a treatment for schizophrenia, whose market values were even greater than that of Xigris. As we went to press in October 2001, the Food & Drug Administration had not yet approved Xigris for sale. In measuring Lilly's investment in developing new products, it is not obvious that failures should be written off, since the successful few were expected to make up for these losses. Certainly Lilly's intangible assets are greater than its total R&D investments. And on average, accountants have found that R&D expenditures result in future profits that justify these investments (see the articles by Dennis Chambers, Ross Jennings, and Robert Thompson.; Baruch Lev and Theodore Sougiannis; and Doron Nissim and Jacob Thomas).

input will almost certainly result in a salable product. As production of intangible assets becomes a more important part of the U.S. economy, this tight, contemporaneous relationship between input and output weakens. Whether any given input will lead to a salable output becomes difficult to predict for individual firms.[14]

Intangible Investments: Hard to Measure, But Not Impossible. Another important difference between tangible and intangible investing is that the firm that makes tangible capital goods is typically different from the firm that will use them. For example, the firm that will use — that is, invest in — computers will generally buy them from another firm rather than making them itself. This makes the investment highly visible: a transaction has occurred, and money has changed hands to attest to the investment's value.

_____

By contrast, intangible investment is generally done in-house: Intel's chips are designed by its engineers, Microsoft's software is designed by its programmers, and Eli Lilly's drugs are developed by its biochemists. So the outlay made to create intangibles is harder to verify. Moreover, while some expenses are clearly aimed at creating intangible assets, other expenses are harder to determine. For example, it is difficult to know how much of a chief executive's time is devoted to producing intangibles and how much to coordinating production.

But it is not impossible. Some corporations attempt to allocate expenses to current production or to future projects. Such corporations require their employees to report work hours on a project-by-project basis. These projects can be classified into those that contribute to current production and those that produce intangible assets. Thus, it might be possible for a corporation to divide money spent on sales and general and administrative needs into expenses for current production and intangible asset production. Doing so might well provide a corporation with a measure of the resources that go into intangible investment that would be of substantial value to its shareholders. If this practice became widespread, statistical analysis would then be possible to evaluate

which proportion of these expenditures result in the creation of an intangible asset.

There are cases in which the intangible investment yields a salable asset. When Chrysler designs a new car, or Eli Lilly develops a new drug, or J.K. Rowling writes a new Harry Potter novel, the design, or the drug, or the novel is a product that could be sold to the highest bidder for a fixed sum. Indeed, this sometimes happens. A design firm such as Pininfarina can design a car for a manufacturer; a small biotech start-up may sell a new drug to a major pharmaceutical company; and a writer may be commissioned to ghost write a book. In these cases, there is no real problem in classifying each of the sales as either income or output.

But with intangible assets it's more difficult. Most of the time, there is no direct transaction to tell us what the intangible asset is worth. Transactions that do tell us about the value of intangible assets are capital transactions: the buying and selling of the equity shares of firms that have invested in and produced the intangible assets. So our only way to measure the success of the vast majority of investment in intangible assets is changes in the stock-market value of firms — which are highly volatile.

MEASURING INCOME AND OUTPUT THROUGH INPUTS AND OUTCOMES

Are there practical ways to measure the major _inputs_ that go into producing intangible assets? If there are, and if most of our investment outcomes are the result of such inputs, we will, over the long run, account for most wealth creation without the sharp ups and downs of the stock market overly influencing our statistics. We do have reasonably good measures of investment in R&D, advertising, and software. But the discussion in this section underscores the difficulties in measuring production of most intangible assets, and the

estimates noted are generally conservative estimates of investments in intangibles.

Consider the various input costs that go into making a new good available to consumers. In the case of a prescription drug, a disease must first be targeted, and an approach to its control or cure must be established. Then a chemical compound must be discovered or constructed that effects the required control or cure. Next, the chemical compound must undergo animal trials, then human clinical trials. Initial clinical trials establish that the compound is safe

domestic product of nonfinancial corporations and 1.8 percent of total GDP.[15] By contrast, in 1978, such corporate R&D expenditures were 1.8 percent of nonfinancial corporate GDP and 1 percent of aggregate GDP. Both of these figures probably underestimate R&D expenditures. Firms that invest in R&D typically have to make additional expenditures to support product development, including marketing, consumer testing, and executive decision making, that are not part of the engineering and scientific expenses that account for most of what the National

agencies; they do not include many other marketing expenses that firms incur, such as the sales forces of pharmaceutical companies or fees paid to public relations firms and athletes — marketing expenses that have been rising faster than agency fees. To the extent that firms spend this money to inform consumers about new products, advertising and marketing expenditures should be counted as investments in intangible assets because the information supplied to consumers through these avenues will generate profits over a sustained period.

Software. One area in which the national income accounts have come to grips with measuring investment in intangibles is software. According to the BEA, in 2000, private businesses invested $183 billion in software, or 1.8 percent of GDP, compared with 0.3 percent in 1978. This software investment comes in three types: prepackaged software; custom software; and own-account software.

Prepackaged software ($61.4 billion in 2000) is sold at arm's length, that is, the company that invests in the software is different from the company that makes it. Sales of prepackaged software to *consumers* have always been counted as consumer expenditures. But such sales to *firms* were counted as expenses, not investment, until the BEA changed its method in 1998. Note that as part of the investment in new software, firms must also train their employees in the use of the software.[16] Thus, purchases of software underestimate the total resources firms must allocate when they invest in new software.

The software investments of firms that purchase prepackaged software do not include the intangible investments made by the producers of

> [Prepackaged software] sales to *firms* were counted as expenses, not investment, until the BEA changed its method in 1998.

and effective. A third round of clinical trials involving large numbers of patients and doctors must determine the range of symptoms for which the drug is effective and the appropriate dosages. These data must be presented to the Food and Drug Administration for approval; a process for mass production for the compound must be designed; and teams of sales personnel must instruct doctors and nurses around the world in the use of the compound. The company may further directly inform patients through print or broadcast media advertising.

Costs of research and development, administration, marketing, and media advertising all enter into the intangible investment. The firm making these investments must believe that these fixed costs will at least be repaid, on average, by the returns to successful intangible assets.

Research and Development. According to National Science Foundation estimates, in 2000, U.S. corporations spent $181 billion of their own funds on R&D. This expenditure represented 3.3 percent of the gross

Science Foundation calls research and development.

Advertising. According to advertising agency McCann-Erickson, firms spent $233 billion on advertising in 2000. This expenditure represents 2.3 percent of GDP, up from 1.9 percent in 1978. However, McCann-Erickson's data reflect the market for advertising

_____

[15] In addition, the National Science Foundation estimates that governments in the U.S., mainly the federal government, spent $72 billion on research and development in 2000, while universities, colleges, and other nonprofit organizations spent an additional $12 billion. In all, $265 billion is estimated to have been spent on research and development, or roughly 2.6 percent of aggregate U.S. GDP. Expenditures by private industry are counted here because all of this expenditure has as its purpose the creation of private intangible assets. Moreover, the public expenditure on research and development is already included in gross domestic product as part of government expenditures. It is also the case, however, that, increasingly, universities, colleges, and other organizations and individuals take advantage of research sponsored by the federal government or nonprofits to license new product development, thereby creating intangible assets.

_____

[16] This point was emphasized by Shinkyu Yang and Erik Brynjolfson.

the software. A company's investment in creating software is separate from the purchasing company's investment in software. For example, Microsoft's investments in producing the Windows operating system and in the Microsoft Office suite of products are separate from the investments that corporations make when they buy these programs. Microsoft's value as an ongoing concern resides primarily in the intellectual property rights it holds for the software it has created and is separate from the value created by other firms' investments to acquire licenses to use Microsoft Windows and Microsoft Office.

Custom software is also purchased, but like custom clothing, it is uniquely adapted for the buyer ($57 billion in 2000). In some of these cases, the rights to the software are sold to the buyer. In other cases, a substantial proportion of the software rights remain with the software producer. When property rights remain with the producer, custom software sales data may understate the value of the producer's investment.

Own-account software is made by employees of the user ($64 billion in 2000). To measure investment in own-account software, the BEA examines how many programmers are employed at firms that don't sell software and estimates how much of their work goes into developing new software (investment) versus maintenance and repair of existing software (expense). The most recent study of this division, which was published in 1982, found that 62 percent of programmers' time was spent on creating new programs.[17] The BEA estimates that since then, programmers have become more involved in repair

and maintenance. Therefore, the BEA counts 50 percent of programmers' time as new software investment, a figure it describes as underscoring the arbitrariness of such measures.

Other Industries' Data Are Sparser. Expenditures on R&D, advertising, and software do not exhaust, by any means, firms' expenditures on intangibles. For example, most financial corporations do not report their expenditures to develop new products as R&D expenses. Yet financial corporations have been making a large and growing investment in financial innovations, including investment vehicles like derivatives and mutual funds, electronic payment systems, ATMs, and credit and debit cards. They have also invested large sums in customer databases and in customer relationships associated with these new instruments.

Almost no data are collected on financial corporations' expenditures on intangibles.[18] However, financial corporations' noninterest expenditures have been rising rapidly. For example, in 2000, noninterest expenditures for commercial banks were $215.5 billion, or 2.1 percent of GDP, up from 1.6 percent of GDP in 1978. Noninterest expenditures include commercial banks' innovations and marketing expenses, but they are only an *indicator* of banks' investment in intangibles because they also include expenditures for tellers and bank branches. The market value of financial institutions has recently

averaged more than 20 percent of the market value of nonfinancial corporations, compared with around 11 percent in 1978. If financial corporations spend proportionally as much on R&D as nonfinancial corporations report spending, this would add another $50

During the earlier period of relatively high saving rates, Americans did not become rich, and as measured saving fell during the 1990s, Americans' wealth increased dramatically.

billion to R&D. Commercial banks alone have added more than $50 billion in noninterest expenditures in this same period. And that neglects the innovative expenditures of mutual funds, insurance companies, real estate firms, other depositories, or investment banks.

Writers, artists, and entertainers make additional investments in intangibles, and these investments are not recorded as part of R&D. In 1997, according to the U.S. economic census, the publishing, motion picture, and sound recording industries had a total revenue of $221 billion. Associated with this stream of revenues are investments in creativity and in finding, developing, and publicizing artists and their work.[19]

Much of the investment in movies, television, and other media pays off quickly because it shows up in movie-theater ticket sales or videotape rentals. Other programming costs, such as many television network broadcasts, are paid for by advertising. However, as Richard Caves points out, television series are produced at a loss — the network's payment for first broadcast rights does not cover the production costs of the series. What producers hope for is that the series will run long enough (three to five seasons has usually been

[17] Thus this study comes from the era before the widespread use of personal computers and computer networks.

[18] See the article by Bob Hunt.

[19] See the book by Richard Caves.

the minimum) so that reruns can be profitably syndicated. Syndication will sometimes pay substantially more than the initial broadcast rights. Similarly, a movie series like "Star Wars" can become a multibillion dollar property, since sequels, video games, toys, and clothes based on the series can be sold.

All told, it can be argued that when the inputs that make up intangible investment are measured more accurately, domestic U.S. corporations' investment in intangibles is likely in the range of $700 billion to $1.5 trillion.[20]

STOCK-MARKET CAPITAL GAINS: USING OUTCOMES TO MEASURE INCOME

The official measures of household income include dividend payments but not stock-market capital gains. The measured personal saving rate is low because stock-market capital gains are high and dividends are low. Personal saving in the United States was low throughout the 1990s, but the net worth of Americans increased from $20 trillion to $41 trillion from the end of 1989 to the end of 2000. Adjusting for inflation, this figure represents a real increase, in 1996 dollars, of $14 trillion (from $24 trillion to $38 trillion).[21] During the three decades before 1990, the U.S. personal saving rate (the ratio of personal saving to disposable personal income) averaged 9 percent. From 1952 to 1989, the annual personal saving rate never fell below 6.9 percent (Figure).

By contrast, in the 1990s, the saving rate averaged much less, 6 percent, and fell during the course of the decade, from 7.8 percent in 1990 to 2.4 percent in 1999. In 2000, it was 1 percent.

But during the earlier period of relatively high saving rates, Americans did not become rich, and as measured saving fell during the 1990s, Americans' wealth increased dramatically. This puzzle remains whether we measure savings and wealth in nominal terms or in real terms.[22]

During the 1960s and 1970s,

---

[22] In nominal terms, during the three decades before 1990, the net worth of American households as a proportion of after-tax income actually fell slightly, from 504 percent to 493. So with the lower saving rate of the 1990s, we might have expected a still lower net worth. Instead, net worth rose to 620 percent of after-tax income at the end of 1999, before falling to 579 by the end of 2000. Alternatively, in real terms, net worth, measured in 1996 dollars, rose from $8.4 trillion at the end of 1959 to $23.4 trillion at the end of 1989 — a $15.0 trillion increase over 30 years and a compound annual growth rate of 3.5 percent. By the end of 1999, net worth rose to $38.1 trillion — a $14.7 trillion increase in just 10 years and a compound annual growth rate of 4.8 percent. Thus, whether we compare increases in wealth with nominal incomes or with consumer price inflation, households' wealth grew more rapidly in the 1990s than in previous decades.

stock-market capital gains were 0.4 percent of GDP. During the 1980s they were 3.7 percent of GDP, and in the 1990s, 16.0 percent.[23] If we use these averages over decades to smooth growth, then from the 1970s to 1980s, the nominal and real growth of the economy, including stock-market capital gains, may have been 0.3 percent higher than reported, and from the 1980s to the 1990s, about 1.2 percent higher.[24]
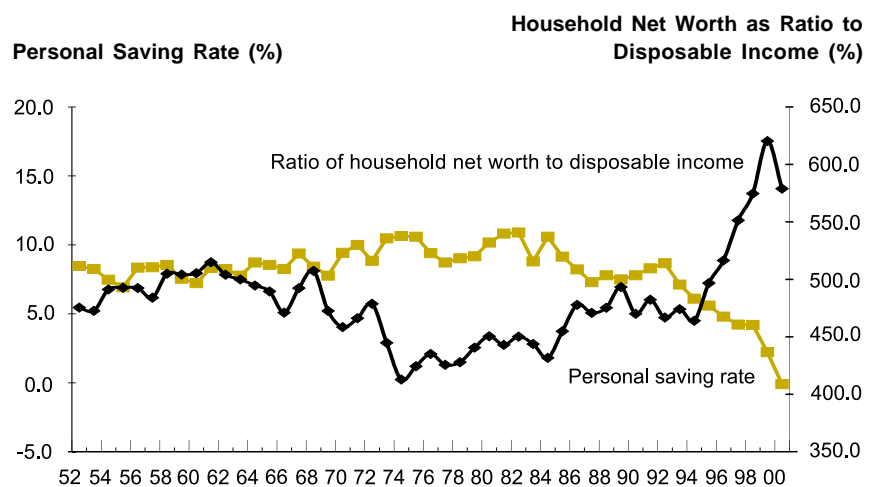
If we attribute this rate of capital gains to intangible investment,

---

[23] From the end of 1959 to the end of 1979, capital gains on equities of domestic corporations, according to the Flow of Funds accounts, averaged just $12.8 billion a year in 1996 dollars, while real GDP averaged $3.6 trillion. From the end of 1979 to the end of 1989, yearly stock-market capital gains averaged $209 billion while real GDP averaged $5.6 trillion. From the end of 1989 to the end of 1999, annual stock-market capital gains averaged $1.2 trillion while real GDP averaged $7.6 trillion.

[24] Thus, if we add capital gains to output, much of the productivity slowdown after the mid-1970s may disappear.

FIGURE

Saving and Wealth

Personal Saving Rate (%)

Household Net Worth as Ratio to Disposable Income (%)

Ratio of household net worth to disposable income

Personal saving rate

---

[20] Further discussion of a variety of data that suggest this is in my working paper.

[21] Specifically, we've used the GDP deflator to eliminate the effects of inflation.

intangible investment must have been quite large. As measured by inputs, investments in intangibles add up to $1 trillion a year.[25] If so, this can help explain why capital gains have been so large.

Some Consequences of Excluding Capital Gains. Excluding capital gains from our measures of household and national income has several disquieting consequences. First, the household saving rate is very low and likely to remain so as long as stock-market capital gains remain strong. Since these capital gains are founded on very large investments in intangible assets, there is little reason to think they will not continue, on average. Of course, volatility will continue, as the recent stock-market downturn reminds us.

Second, if stock options continue to rise in importance as a form of reward to employees, employee compensation will increasingly depend, at least in part, on stock-market capital gains. This compensation can be measured in terms of the market value of the option when issued or in terms of the realized value of the option when it's exercised. How to properly measure this compensation in our accounts is a question that is yet unanswered. At present, most employee stock options are included in personal income when they are exercised, not when they are granted. Recently, personal income for 2000 was revised upward, in part because the amount of stock options exercised was larger than initially anticipated. As a result, measured personal saving rose from a negative to a low positive number.

Third, when stock options are

exercised or when stocks are sold and capital gains are realized, tax obligations are accrued. These capital gains taxes have been an important element of the surge in personal income tax payments in the late 1990s that has continued into the new millennium. As a consequence, tax payments as a proportion of measured household income have risen. Thus, even if we ignore capital gains in our income and compensation measures, they have an important impact on government finance and measured household saving, since increased personal tax payments raise government saving and lower household saving.

Finally, the income of financial intermediaries often feeds off capital gains. For example, firms that manage investment funds often earn a proportion of the capital gains they accrue on behalf of clients, and an investment bank may make a substantial fraction of its income from capital gains. How to include such earnings in the national accounts is not easily determined, but since such corporations account for a fifth of all stock-market equity, they are an important part of the economy.

CONCLUSION

Changes in the U.S. economy have made U.S. economic developments inherently more difficult to analyze. In particular, production becomes riskier as more of our efforts are devoted to producing intangible assets. Measuring this effort is hard, and measuring its outcome is even harder. Yet making the effort to measure these investments is surely preferable to ignoring them, even though the outcome is not entirely satisfactory.

If we were to include increases in households' net worth in GDP, the variability of these capital gains would overwhelm that of the rest of income. In

1999, real household net worth rose by $4 trillion (in 1996 dollars); in 2000 it fell about $2 trillion. Since real GDP was roughly $9 trillion in 1999, real GDP including these capital gains was about $13 trillion; in 2000, it tumbled to $7 trillion.[26] Thus GDP growth measured this way was negative by more than 40 percent! That decline is the amount we would generally associate with an economic catastrophe like the Great Depression. Yet the unemployment rate scarcely changed between 1999 and 2000; in fact, it fell slightly from an average of 4.2 percent to 4.0 percent.

It may well turn out that excluding capital gains from our measures of national income and living with a spuriously low personal saving rate may be the best alternative. However, we might wish to add another measure of household income and saving that does include capital gains. Indeed, we might want to have one measure that includes capital gains that have been realized, that is, where the investor has taken the profit by actually selling the stock, and another one that includes all stock-market capital gains, realized and unrealized.

It may not be possible to use a single standard of GDP as our sole measure of U.S. economic progress. Nevertheless, we should continue to improve our measures of GDP. The BEA has taken an important step by including software investment in GDP. Other items the BEA should consider in the future include R&D and advertising. ⓑⓡ

---

[25] For details, see my forthcoming working paper.

[26] To be more precise, if we use the GDP deflator to convert net worth into 1996 dollars, in 1999 households' net worth rose $4.2 trillion and in 2000 it fell $1.9 trillion. In 1999, real GDP without capital gains was $8.9 trillion, and in 2000 it was $9.2 trillion. Thus, including capital gains, real GDP was $13.1 trillion in 1999, and $7.3 trillion in 2000, a decline of 44 percent.

REFERENCES

Campbell, John Y., Martin Lettau, Burton G. Malkiel, and Yexaio Xu. "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk," *Journal of Finance* 56, 2001, pp. 1-43.

Caves, Richard. *Creative Industries: Contracts Between Art and Commerce.* Cambridge, MA: Harvard University Press, 2000.

Chambers, Dennis, Ross Jennings, and Robert B. Thompson. "Evidence on the Usefulness of Capitalizing and Amortizing Research and Development Costs," mimeo, University of Texas, January 1998.

Fama, Eugene F., and Kenneth French. "Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay?" *Journal of Financial Economics* 60, April 2001, pp. 3-43.

Greenwood, Jeremy, and Boyan Jovanovic. "The IT Revolution and the Stock Market," *AER Papers and Proceedings*, May 1999.

Greenwood, Jeremy, and Mehmet Yorukoglu. "1974," *Carnegie-Rochester Conference Series on Public Policy*, 46, June 1997, pp. 49-95.

Hall, Robert E. "The Stock Market and Capital Accumulation," NBER Working Paper 7180 (June 1999).

Hunt, Robert M. "You Can Patent That? Are Patents on Computer Programs and Business Methods Good for the New Economy?" Federal Reserve Bank of Philadelphia *Business Review* (First Quarter 2001).

Jovanovic, Boyan, and Peter Rousseau. "Vintage Organizational Capital," NYU Working Paper (2001).

Keynes, John Maynard. *The General Theory of Employment, Interest, and Money.* New York: Harcourt Brace, 1938.

Kuznets, Simon. *National Income and Its Composition, 1919-1938, Volume 1.* New York: National Bureau of Economic Research, 1941.

Lev, Baruch, and Theodore Sougiannis. "The Capitalization, Amortization, and Value Relevance of R&D," *Journal of Accounting and Economics* 21, 1996, pp. 107-138.

Nakamura, Leonard. "Intangibles: What Put the *New* in the New Economy?" Federal Reserve Bank of Philadelphia *Business Review* (July/August 1999).

Nakamura, Leonard. "Economics and the New Economy: The Invisible Hand Meets Creative Destruction," Federal Reserve Bank of Philadelphia *Business Review* (July/August 2000).

Nakamura, Leonard, "Education and Training in an Era of Creative Destruction," Federal Reserve Bank of Philadelphia Working Paper 00-13/R (2000).

Nakamura, Leonard. "What Is the U.S. Gross Investment in Intangibles? (At Least) One Trillion Dollars a Year!" Federal Reserve Bank of Philadelphia Working Paper 01-15 (2001).

Nissim, Doron, and Jacob Thomas. "R&D Costs and Accounting Profits," University of Haifa, Zimmerman Foundation Discussion Paper ZF-01-01, April 2000.

Peach, Richard, and Charles Steindel. "A Nation of Spendthrifts? An Analysis of Trends in Personal and Gross Saving," Federal Reserve Bank of New York *Current Issues in Economics and Finance* 6 (September 2000).

Romer, Paul. "New Goods, Old Theory, and the Welfare Costs of Trade Restrictions," *Journal of Development Economics* 43, 1994, pp. 5-38.

Rudebusch, Glenn. "New Estimates of the Recent Growth in Potential Output," Federal Reserve Bank of San Francisco *Weekly Letter* 95-40, Nov. 24, 1995.

Scherer, Frederic M., and Dietmar Harhoff. "Technology Policy for a World of Skew-Distributed Outcomes," *Research Policy* 29, 2000, pp. 559-66.

Schumpeter, Joseph. *Capitalism, Socialism, and Democracy.* New York: Harper, 1942.

Solow, Robert M. "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics* 70, 1956, pp. 65-94.

U.S. Bureau of Economic Analysis. *Private Nonresidential Fixed Investment in Software* (seasonally adjusted).

Yang, Shinkyu, and Erik Brynjolfsson. "Intangible Assets and Growth Accounting: Evidence from Computer Investments," paper presented at New York University's Fourth Intangibles Conference, 2001.