

New Beginnings

BY ANTHONY M. SANTOMERO

Welcome to the first issue of the *Business Review* in its new format. Our regular readers will notice right away that the new *Business Review* is bigger. In part, the larger size reflects the fact that we've gone to a quarterly schedule (four issues a year), rather than the previous bimonthly schedule (six issues a year).

As you read through, you'll see it also has more to offer. We continue to bring you our economists' analysis of current issues surrounding the economy, banking, and the financial sector. But we will also share some of the insights that flow from their long-term research projects, and we will report on conferences and seminars held at the Bank and around the District.

One new feature of the *Business Review* I am particularly happy about is this column, *The Third Dimension*, where I will have the opportunity to share some of my thoughts with you. In future columns I plan to write about monetary policy and other central banking issues. This time, though, let me simply introduce myself and say a few words about the region in which our Bank operates.

I came to the Federal Reserve Bank of Philadelphia as its ninth president in July 2000. Prior to my appointment, I had been professor of finance at the Wharton School of the University of Pennsylvania and director of the Wharton Financial Institutions Center. I have lived in the Philadelphia area since joining the Wharton faculty in 1972. Over the years, academic conferences and consultations

with bankers have taken me to many different parts of the world, but Philadelphia is home to me. It is a pleasure to be at the helm of one of its pre-eminent institutions.

I believe that a great strength of the Federal Reserve System is its network of Reserve Banks that weaves central banking into the fabric of the nation's diverse regional economies. Our Bank serves the Fed's Third District – eastern Pennsylvania, southern New Jersey, and Delaware. We oversee the banking organizations that operate here. We provide depository institutions here with coin and currency, clear their checks, and move funds electronically on their behalf. Most important, we represent the people and businesses of the District in the nation's monetary policy deliberations and decisions.

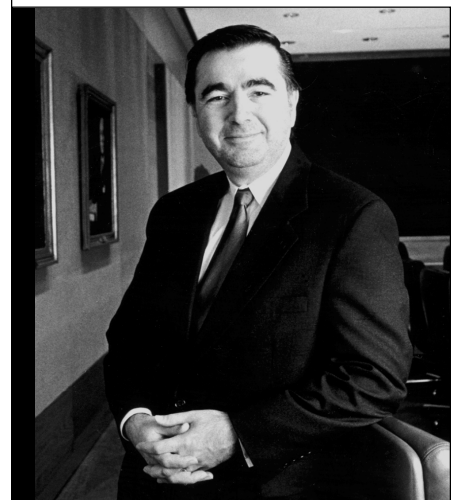
We know that to do our job well, we must stay in constant contact with the banks, businesses, and people of the District. We strive to bring the personal touch to the services we provide. We talk with District leaders regularly, both formally through our board of directors and advisory councils, and informally in the everyday course of conducting the Bank's busi-

ness. We study District economic and financial data to understand the trends at work here.

Recently, our Research Department put together an overall assessment of economic conditions and prospects of the largest metropolitan area in the District, Philadelphia. I had the opportunity to present our findings to the Greater Philadelphia Chamber of Commerce, and it generated a good bit of interest and discussion. The complete report is available at www.phil.frb.org/files/reghigh/repcard01.pdf. I'll just share some of the highlights with you here.

Philadelphia's Economic Performance and Prospects

Philadelphia is part of a metropolitan area of 5 million people spanning nine counties across Pennsylvania and New Jersey. Over the past decade it has shared in the national economic expansion in many ways. The region's unemployment rate has declined dramatically. Venture capital flowing into the area increased sixfold during the late 1990s. In the high-tech sector, Philadelphia is among the nation's leaders for investments in pharmaceuticals and biotechnology. Sustained



Anthony M. Santomero, President,
Federal Reserve Bank of Philadelphia

economic growth and effective financial management have helped local municipalities achieve sound financial conditions. For example, the city's fiscal situation has greatly improved over the past several years.

Despite these successes, metropolitan Philadelphia's overall growth has been relatively slow. Although population in the United States grew more than 10 percent in the 1990s, and cities like New York and Boston had some increase in population, the Greater Philadelphia area has seen little change in its total population. Indeed, Philadelphia ranked 42nd out of 50 metro areas in job growth during this expansion.

The slow growth partly reflects the general movement of people, jobs, and economic activity away from the old, established Northeast urban centers and toward the South and West. But other factors are slowing growth here, too. In general, higher labor costs (even adjusted for the quality of our workforce) and high energy costs have dampened job growth in the region. The high tax burden in the city of Philadelphia has also deterred job creation. These factors would have a substantial impact on the attractiveness of any community and so affect growth and job creation here. The good news is that many of these factors are ones that Philadelphia can reverse and turn to its advantage.

In today's knowledge-based economy, the key to a region's success is to attract and retain a growing pool of well-educated and highly skilled people. Philadelphia has not done so well in that regard. For instance, the region ranks relatively low in the proportion of its workforce with college degrees. But it can improve this situation in a number of ways.

Philadelphia is, in fact, home to many highly rated colleges and universities. Clearly, it needs to prepare and encourage more young Philadelphians to go on to college and earn their degrees. Improving the quality of the region's public primary and secondary schools is among the necessary steps in that process.

Also, because of their reputation, local colleges and universities

attract the best and the brightest students from across the country and around the world. Regional employers should pursue these students aggressively when they graduate in order to keep them in our region.

Attracting more skilled workers from abroad offers another opportunity to increase the regional workforce, and one that Philadelphia has not fully exploited. For instance, why did the populations of Boston and New York grow during the 1990s while Philadelphia's did not? One important reason is that those two cities attracted their share of new immigrants to the U.S. while Philadelphia did not. Perhaps our region's trade missions to attract foreign capital should focus on attracting foreign labor as well.

Developing and attracting talented and educated people is important. Keeping them here is equally important — and equally challenging. Good people are mobile, and they choose to make their homes in places where they can enjoy life.

Providing a rich quality of life is perhaps Philadelphia's strongest suit. Philadelphia is steeped in American history. Consider this: Across the street from our Bank, a national Constitution Center is being built to complement Independence Hall and the Liberty Bell, both of which are nearby. In the process of excavating the site for the center, construction crews came upon layers of earth preserving literally 1 million artifacts belonging to African, European, and Native American people living in the area as far back as 1650.

The city is home to an array of major cultural institutions from its world famous orchestra to its art museum. It boasts a growing list of fine restaurants too numerous to mention and a full calendar of events ranging from the Mummers parade on New Year's Day to the annual bicycle race through the hills of Manayunk. Within a few hours' drive, Philadelphians can be on an ocean beach, at a mountain ski resort, or in a casino.

For a long time, Philadelphia's amenities were not well publicized and, hence, not well known. That seems to be changing. For instance,

the city received high marks for the way it hosted the Republican National Convention last summer.

But perhaps Philadelphia's biggest challenge is alleviating the heavy tax burden on its residents. Philadelphia residents bear one of the highest tax burdens of large-city residents anywhere in the country. Everyone recognizes that reducing the city wage tax is crucial to making the city a more attractive place, and the city has taken some initial steps in that direction. The challenge is to make significant cuts in tax rates without either endangering the city's fiscal health or compromising on the goal of improving the quality of life here. Providing good schools, safe streets, attractive public spaces, and an efficient transportation network — all these things cost money. Yet all are essential to attracting and retaining good people.

Thus, finding creative ways to improve government's value proposition — that is, for government to provide better basic services at lower tax rates — is crucial to the region's future success.

In the end, the results of our economic "report card" for Philadelphia drove home two points for me. The first is that for Philadelphia to do well — indeed for any region in our District to do well — we need sustained growth in the national economy.

The second is that sustained growth in the national economy is not enough. For Philadelphia or any other region to reach its full potential also requires the concentrated efforts of its businesses, governments, educational institutions, and other organizations to make that region a location of choice for the economy's most productive people.

I believe that the Philadelphia Fed has something to contribute on both counts. We will certainly do our best.

We welcome your feedback on this and future issues of the *Business Review*. Please e-mail us at Phil-BRComments@phil.frb.org.

You Can Patent That?

Are Patents on Computer Programs and Business Methods Good for the New Economy?

BY ROBERT M. HUNT

The United States is in the midst of an economic boom sustained in part by rapid technological innovation.

Firms are always looking for ways to enhance or protect their market position.

Increasingly, they are turning to patents to protect not just physical inventions, but more abstract ones such as computer programs and even ways of doing business. Two decades ago, many of these patents would have been impossible to obtain, let alone enforce. But almost every day now, another example of these patents is described in the press.

Are these patents really a new phenomenon? Are they good for the economy? This article describes how changes in patent law made it possible for inventors to obtain patents on discoveries as abstract as lines of computer code or simply a way of conducting business. It also examines

the potential economic benefits and costs of these new patents.

While it is too soon to quantify these effects, there are good reasons for concern and a number of things we can do to address those concerns. At a minimum, we must reserve patents for inventions that represent more than an obvious combination of existing technologies. We should be willing to increase the resources and expertise available to the patent office to ensure it knows what has already been invented. And when patent disputes reach the courts, decisions of the patent office should carry no more weight than is warranted by the quality of its examinations. Over time, we may find that more radical measures are required. Or we may find that the patent system will adapt to this latest in a series of technological revolutions.

ENGINES OF GROWTH

In our economy, rising labor productivity is a prerequisite for sustainable economic growth and

rising incomes. Between 1996 and 1999, labor productivity grew at an average rate of 2.6 percent a year, much faster than the 1.6 percent average annual increase experienced over the previous two decades.¹ This allowed the U.S. economy to grow more rapidly, and with less inflation, than most economists dared to predict five years ago.

Economists Stephen Oliner and Daniel Sichel estimate that at least two-thirds of the increase in the growth rate of labor productivity since 1995 is due to heavy investment in computer hardware, software, and communication technologies, and the rapid improvement of those technologies. Their study shows that, since 1990, investment in computer software alone has contributed more to growth in labor productivity (in fact twice as much) than all investment in capital excluding information technology.

Advances in computer software, combined with large investments in computing and communications hardware, are enabling the rapid expansion of an entirely new medium of commerce — the Internet. While e-commerce accounts for only a small share of economic activity today, it is growing many times faster than the bricks-and-mortar economy. One of its most important applications is in financial services. Already a significant volume of securities trading is initiated by orders placed at a web site. Internet banking and electronic bill payment are predicted to grow rapidly over the next decade.²

¹ This calculation is based on the Bureau of Labor Statistics index of labor productivity in the nonfarm business sector of the economy for the periods 1976-95 and 1996-99.



Bob Hunt is an economist in the Research Department of the Philadelphia Fed.

Competition in this new medium is intense. Naturally, companies are looking for ways to enhance or protect their market position. Increasingly, they are turning to patents to protect their investments in computer software and even their business models. According to the U.S. Patent and Trademark Office, fewer than 100 Internet-related patents were issued prior to 1992. Over the next five years, the patent office granted 750 Internet-related patents. In 1999 alone, it granted nearly 4000 Internet-related patents.³ Only a decade ago virtually none of these patents would have been granted, and few companies would have bothered to file an application. What happened?

PATENTS ARE NOT FOR ALL INVENTIONS

The American patent system is designed to reward inventors that make new and useful discoveries (see *Patent Basics*). But our patent laws include certain limitations that preclude the patenting of some inventions.⁴

² So far, growth in these services has been disappointing. In an article in the *Business Review*, Loretta Mester examined a number of factors that explain why adoption of these new payment methods is likely to be initially slow. These include large up-front costs for deployers; network effects, which limit consumer interest in a payment method until its widespread adoption; and the balance of risk and benefits of a new payment method as compared to existing ones.

³ Computer and telecommunication equipment manufacturers and software developers own most of these patents. But here are just a few of the interesting exceptions to the rule: Walker Asset Management (owners of the Priceline patent and 24 others), Citibank (15), Incyte Pharmaceuticals (the firm responsible for mapping a significant part of the human genome, 9 Internet patents), Merrill Lynch (9), Amazon.com (7), VISA International (7), Chase Manhattan (6), Andersen Consulting (5), E-Stamp Corp. (4), and Cybercash (one of the first developers of e-cash, with 3 patents).

⁴ The U.S. law of patents can be found in Title 35 of the U.S. Code (<http://uscode.house.gov/uscode.htm>), which incorporates all of the various patent statutes enacted over the years.

Patent Basics

F

or over 200 years, the U.S. government has used patents to reward inventors for their discoveries.^a The reward is a grant of the legal right to exclude others from making, using, or selling the patented invention for a limited period of time. But monopolies imply higher prices and, thus, less consumption of patented products or processes, which is socially wasteful. A golden rule, then, is that patents should be granted only for discoveries that are really new.

While economists favor the establishment of well-defined property rights, they also recognize that doing so entails some cost—at a minimum, the cost of resolving any disputes that may arise. That is one reason most countries go a little further, granting patents only for inventions that are not trivial extensions of what is already known (the legal concept is called nonobviousness). What's more, the monopoly granted ought to cover only what the inventor discovered and no more. But to do so, the inventor must describe the invention in sufficiently precise terms. If that description is made available to the public, granting patents can assist in the more rapid dissemination of technological knowledge.

Each of these features is embodied in the American patent system, which involves inventors, specially trained attorneys, the U.S. Patent and Trademark Office, and the federal courts. An inventor typically obtains the services of a patent attorney to undertake the process of applying to the patent office for a patent on his or her invention. The patent application will contain a description of the invention, a discussion of any related inventions or techniques known to the inventor (what is called the *prior art*), and a set of proposed claims that will define the property rights he or she is seeking. Examiners at the patent office review the application, conduct their own search of the prior art, decide whether to grant a patent, and if so, specify the precise language of the patent's claims. In essence, the inventor and the patent office negotiate a custom-designed property right tailored to reflect what was invented.^b But a patent can be granted only when an invention satisfies the requirements set out in the patent statute, as interpreted by federal courts.

An inventor who disagrees with a patent examiner's decision may choose to refile the application and have it reconsidered, or appeal the decision to an administrative panel and, from there, to a federal appeals court. If a patent is granted, and the owner feels the patent is being infringed, he or she can sue the offending party in a federal court. The patent holder can seek a court order prohibiting any further infringement (an injunction) and compensation in the form of lost profits or a reasonable royalty. Defendants in a patent suit typically argue they did not infringe the patent and that the patent is invalid, that is, the patent office made a mistake when it decided to grant the patent. The federal courts decide these cases on the basis of their precedents and the language of the patent statute. Some infringement cases are settled before a verdict is reached. This typically involves a licensing agreement whereby the defendant agrees to pay for the right to use the invention or to license some of its own patented technologies to the other party.

^a The first U.S. patent was granted in 1790 to Samuel Hopkins of Philadelphia, who invented a new way of making potash (used in soap, glass, and gun powder).

^b The patent office's online database of patents can be found at its web site: <http://www.uspto.gov>.

Patentable Subject Matter.

Assuming the criteria described in the next section are also satisfied, any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement of these things, can be patented. These categories are quite broad, but the courts

have identified certain types of subject matter that cannot be patented, including laws of nature, physical phenomena, and abstract ideas. Prior to 1980, most patent attorneys believed these exceptions precluded the possibility of patenting computer software or methods of doing business.

Patentability Criteria. To qualify for patent protection, an invention must satisfy the requirements of *utility*, *novelty*, and *nonobviousness*. Utility simply means that the invention is useful. Novelty means the invention is truly something new. An invention fails the test of novelty if before the inventor applied for a patent something very much like it already existed or was described in print. Existing products or processes, existing patents, or articles on the subject in technical publications are called the “prior art.”

The requirement of nonobviousness goes a bit further than the requirement of novelty. A patentable invention must be something more significant than a trivial extension of the prior art. In hindsight, an invention may seem pretty obvious, but that is not the standard used in patent law. Patent law asks, would the invention have been obvious, at the time it was made, to a person with ordinary skill in the field and with knowledge of the relevant prior art? If the answer is yes, the invention is obvious and a patent will not be granted.⁵

CAN I PATENT THAT? WELL, NOW YOU CAN

Over the last two decades, the American patent system has changed in many ways. The definition of patentable subject matter has gradually been expanded, reversing the traditional view that computer programs and methods of doing business were unpatentable. In addition, patentability criteria have become less strict. These trends arose from a series of court decisions beginning in the early 1980s. Their effect has been amplified by certain other changes, described below. The result is a patent system that operates very differently than it did 20 years ago.

⁵ A more detailed discussion of the nonobviousness requirement can be found in my previous article in the *Business Review*.

The Early Years. In the 1960s and early 1970s, computer programs enjoyed very limited intellectual property protection. And it was widely believed that patent protection for computer programs was unavailable. This impression was bolstered by a 1972 Supreme Court decision (*Gottschalk v. Benson*) involving an application to patent a computer program that translated decimal numbers into binary numbers. The court concluded the program was a

In the 1960s and early 1970s, computer programs enjoyed very limited intellectual property protection.

mathematical algorithm, which, like laws of nature or an abstract idea, does not fall into one of the categories of patentable subject matter.

This lack of protection did not matter as much then as it does today. At the time, most computers and computer programs were custom-designed. As a result, most computer programs would not work on another machine without significant modification. Also, companies developed more of their own computer programs and limited outsiders’ access to their code. In this environment, firms could protect their programs as *trade secrets*, using state law to prosecute those who stole the code or disclosed it to the public.⁶

When the computer industry began moving away from custom-designed machines, toward standardization and mass production, more

⁶ The key to asserting trade secret protection is demonstrating that substantial precautions were taken to prevent the secret from being disclosed. See the article by Friedman, Landes, and Posner. Since that article was written, the Economic Espionage Act of 1996 (18 U.S.C. 831-9) created new federal criminal penalties for certain instances of misappropriation of trade secrets.

computer programs would run on more machines. Firms began to purchase more software “off the shelf,” and more firms began to develop software with the intention of selling it to as many customers as possible. But when a company widely distributes a computer program, protecting it as a trade secret becomes more difficult.

As trade secret protection became less effective, a number of firms began to seek other legal protections. After many years of study and debate, Congress modified the Copyright Act in 1980 to explicitly extend copyright protection to computer programs. But by the early 1990s, it was clear that copyright afforded relatively narrow protection for software, allowing rivals to offer very similar products without infringing the copyright. Some firms sought broader forms of protection from imitation. What they eventually got was patent protection for computer programs.

Computer Programs Become Patentable Subject Matter.

Long before the meaning of copyright protection for computer programs was well defined, the Supreme Court opened the door to the possibility that computer programs could be patented.⁷ In a 1981 decision (*Diamond v. Diehr*), it ruled that an invention using temperature sensors and a computer program to calculate the correct curing time in an otherwise conventional process of molding rubber goods could be patented. The patent office had rejected the patent application, arguing the only new aspect of the invention was the computer program, which repeatedly solved a well-known chemical equation using the temperature data provided by the sensors. The Supreme Court disagreed, arguing that the invention was an improved process for making rubber goods that happened to use a computer program. Thus, the

⁷ The Supreme Court later argued it had never ruled that all computer programs were unpatentable, only that the computer programs involved in the early cases were not patentable.

court seemed to distinguish between mathematical algorithms per se and the application of an algorithm to accomplish something useful, concrete, or tangible.⁸ Patent attorneys learned to write software patent claims to emphasize the idea of physical transformations that produce useful, tangible results. They also began to write patent claims in terms of a computer program embedded in a machine (that is, a digital computer). Over time, both approaches gained acceptance at the patent office and in the courts.

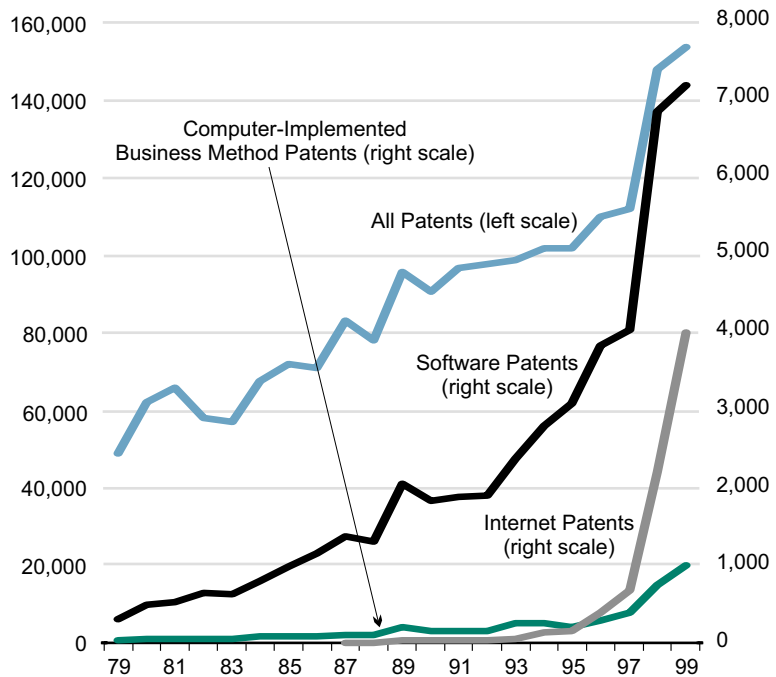
More recent court decisions place more emphasis on the importance of generating useful results than on producing concrete or tangible results. The link to physical transformations has also become less important. A 1994 decision (in *re Alappat*) upheld the patentability of a computer program that smoothes digital data before displaying it as a waveform on a computer monitor. The patent office rejected the application, pointing out the lack of any physical transformation of matter. The court disagreed, arguing the invention, a computer programmed to compute averages in a particular way, was a machine and therefore patentable subject matter. It now appears that only computer programs producing the most abstract outcomes are at risk of falling outside the categories of patentable subject matter.⁹

Inventors clearly noticed this change, and the number of software patents issued increased dramatically

after 1994 (Figure), much more so than the overall increase in patents granted. Many of the firms that are most prolific in patenting computer programs are not software companies per se. Rather, they are the leading manufacturers of electronics and computing devices.

What About Methods of Doing Business? Until very recently, the conventional wisdom was that business methods did not fall into the categories of patentable subject matter. But this conclusion was never as absolute as it is frequently described, especially in the popular press. The

FIGURE
Patents Granted in the U.S.



Source: U.S. Patent and Trademark Office and author's calculations.

Notes: The chart uses data reported for calendar years. *All patents* are all patents on inventions. This count excludes design and plant patents. The remaining categories are not mutually exclusive, nor are they very precise. The category *Internet* is based on a keyword search conducted by the patent office in July 2000. The category *computer-implemented business methods* is based on the author's search for patents falling into patent classification 705 (data processing: financial, business practice, management, or cost/price determination) conducted in October 2000. The category *software* is based on a count of patents falling into certain classifications (364, 395, 700-7, 713-4, and 716-17), according to patent office statistics released in April 2000.

The exact number of software patents in existence is a subject of considerable debate. The U.S. patent classification system does not divide inventions into those that are exclusively software and those that are not. The ubiquity of computer programs and the frequency with which software is closely integrated with specific hardware make such a distinction difficult to implement, even if it were not misleading. The count provided above is based on the patent classes that clearly account for many types and uses of software. But some of these classes also include inventions we would typically describe as hardware. And software programs classified elsewhere are excluded from this count.

According to the approach used by the author, the patent office has granted roughly 50,000 software patents since 1978. A keyword search for the terms *computer program* or *software* would turn up about 120,000 patents issued since the early 1970s. It is likely that the count shown in the figure exaggerates the number of software patents issued in earlier years while missing many more issued in recent years.

⁸ This distinction is not new. A century before computer programs, American courts concluded that new ways of making a machine, manufacture, or composition of matter (that is, processes) were indeed patentable. The distinction – between a patentable process versus the law of nature, or abstract idea, that makes the process work – often turned on whether the process involved a physical transformation of matter.

⁹ In recent law review articles, John Thomas and Arti Rai have argued that the current method of determining if a computer program qualifies as patentable subject matter has been reduced to a single question: "Is the program useful?"

patent office, for example, recently identified 41 instances of what it calls financial patents issued *before 1848*.¹⁰

The modern variety of this kind of patent usually involves a computer. An early example of what the patent office now calls “computer-implemented business methods” is Merrill Lynch’s development of the cash management account in the 1970s. This system involved three accounts linked through a computer program. In 1982, Merrill Lynch obtained a patent on the computer system and the software to implement its cash management account. Rival Paine Webber sued to invalidate the patent, arguing that the invention was an unpatentable business method. But a federal court disagreed, concluding that Merrill Lynch’s invention “...effectuates a highly useful business method and would be unpatentable if done by hand.”

By the early 1990s, the patentability of computer software was clearly established. What’s more, development of commercial applications of the Internet exploded. Suddenly, everyone was developing a business model, which typically involved using computers and software to conduct old forms of business in new ways. So it is hardly surprising to find that the number of applications for computer-implemented business methods has grown rapidly in the last few years, and the patent office is granting more and more of these patents (see *A Few Patents Involving Computer Software or Business Methods*).

Any doubt about the viability of patenting a business method that relies on computer hardware and software was eliminated in a single court decision in 1998 (*State Street v. Signature Financial Group*). In that

¹⁰ That number is cited in the recent USPTO White Paper. It includes two patents on bank notes, two patents on financial instruments, four patents on checks, and five patents on interest-calculation tables. The White Paper also cites a patent, issued in 1907, for an insurance system.

decision, the court doubted that a precedent establishing an exception for business methods had ever existed. Even if there was one, the court concluded it was irrelevant under current patent law. The patentability of a business method, according to this

Twenty years ago, it was not uncommon to see patents invalidated by federal courts.

decision, depends on whether the claimed invention is useful, new, and nonobvious.

But Patentability Criteria Were Also Relaxed. Twenty years ago, it was not uncommon to see patents invalidated by federal courts, often because the claimed invention was found to be obvious in light of the prior art. There was also at least the appearance that patent cases were being decided differently in different parts of the country. Some policymakers felt the U.S. was losing ground to other countries in certain high technology industries and the patent system was to blame. Each of these concerns contributed to the decision in 1982 to create a new court (the Federal Circuit) to hear all appeals of patent and certain other cases. It was hoped that a single appeals court would reduce any unequal treatment of patent cases in the different district courts and that, by hearing patent cases more regularly, the court would be able to develop greater expertise in a highly specialized area of law.

The decisions of this new court soon altered the landscape of U.S. patent law.¹¹ The most significant change was the modification of the test for determining the obviousness of a claimed invention. The new court was more willing to rely on secondary factors, such as evidence of

¹¹ For additional detail on these changes, see my 1999 Working Paper and my previous article in the *Business Review*.

commercial success, to indicate that an invention was nonobvious. In practice, the new test is much easier to satisfy than the one used over the previous quarter century and, as a result, many more inventions now qualify for patent protection.¹²

Other Important Changes.

In another change, the new appeals court strengthened the presumption that the patent office was correct in issuing a patent.¹³ As a result, more evidence is required to invalidate a patent. Other decisions made it easier for a patent holder to obtain a preliminary injunction — a court order prohibiting a potentially infringing activity before the question of infringement is definitively decided by the court.¹⁴ If we compare trial outcomes before and after 1982, we find that courts are much less likely to invalidate patents and more willing to issue preliminary injunctions.¹⁵

¹² Were these changes important? In 1994 Ronald Coolley argued that “many patent attorneys believe that the obviousness defense is dead and that the cause of death lies in the decisions of the Court of Appeals for the Federal Circuit.” In 1991, Lawrence Kastriener claimed that “as a result of these changes, patents today are more likely to be held valid than, perhaps, at any time in our history.” The title to his article, “The Revival of Confidence in the Patent System,” conveys the sense among some observers that a broken system had been repaired by the decisions of the Federal Circuit. Other observers would disagree. See, for example, John Barton’s article.

¹³ Prior to 1986, federal courts frequently decided the validity of a patent based on whichever side presented more convincing evidence (*a preponderance of the evidence*). In 1986, the Federal Circuit ruled that a patent is presumed to be valid until proven otherwise by *clear and convincing evidence*, a more difficult standard to satisfy. See *Medtronics Inc. v. Intermedics, Inc.* and *Hybritech Inc. v. Monoclonal Antibodies Inc.*

¹⁴ Before 1982, federal courts would not grant a preliminary injunction if they had any reasonable doubt about the validity of the patent in question. The Federal Circuit relaxed this evidentiary standard.

¹⁵ The articles by Adam Jaffe, M.A. Cunningham, and Donald Dunner and his colleagues review quantitative evidence of the change in trial outcomes.

A Few Patents Involving Computer Software or Business Methods

One-Click Purchasing on the Internet

In 1999, Amazon.com obtained a patent (no. 5,960,411) for a computer program that stores a customer's address and credit card number in a database; the program allows the customer to make a purchase with a single mouse click. Amazon's rival, Barnes and Noble, implemented a similar system at its web store. In December 1999, Amazon obtained a preliminary injunction, preventing Barnes and Noble from using a one-click ordering process on its web site. Barnes and Noble added an extra click to its ordering system and had the injunction thrown out on appeal in January 2000.

"Name Your Own Price" Purchasing on the Internet

In 1998, Walker Digital, Inc. obtained a patent (no. 5,794,207) on a computer system and software that enable reverse auctions over a communications network. This is the most famous of the many computer-implemented business methods patented by Walker Digital. Why? Because it is the patent behind Priceline.com's "name your own price" reverse auction system for selling airline tickets, hotel rooms, and car rentals. Soon after obtaining the patent, Priceline sued Microsoft's Expedia travel service for infringement. Another company, Marketel International, sued Priceline, alleging the patented technology was developed by its employees and not Walker Digital. In January 2001, Priceline and Expedia reached a licensing agreement.

Delivering Music Over a Communications Network

InTouch Group owns patents (nos. 5,237,157 and 5,963,916) on a method of delivering portions of pre-recorded music at kiosks and over the Internet. It recently sued Amazon.com and three other companies that allow customers to sample songs contained in CDs for sale on their web sites.

The Cash Management Account

In 1982, Merrill Lynch received a patent (no. 4,346,442) on a computer system and software that enabled financial transactions for the cash management accounts it offered to investors. It was actually a set of three accounts, which included features typically associated with a checking account. Unlike most patents of this sort, it has been tested by litigation. Paine Webber sued Merrill Lynch to invalidate the patent on the grounds that the computer program was an unpatentable algorithm and the cash management system was an unpatentable business method. A federal district court rejected both of those arguments.

The *CollegSure* CD

In 1989, New Jersey's College Savings Bank obtained a patent (no. 4,839,804) on its special certificate of deposit, which pays a return commensurate with the increase in the cost of college tuition. Shortly after obtaining the patent, the bank sued CenTrust Savings Bank for infringement. The case was settled prior to trial after CenTrust agreed to pay licensing fees to College Savings Bank.

A Data Processing System for Managing Mutual Funds

In 1993, Signature Financial Group obtained a patent (no. 5,193,056) on a data processing system that enabled a "hub and spoke system" for mutual funds. The system allowed a fund manager to aggregate the assets of several mutual funds into a single portfolio, reducing overhead costs while maintaining the necessary transaction information for allocating gains, losses, and tax liabilities to the original mutual funds. State Street Bank and Trust sued Signature Financial to invalidate the patent. It succeeded in the original trial, where the judge concluded that Signature's computer program was an unpatentable algorithm as well as an unpatentable business method. Both of those conclusions were decisively reversed on appeal in 1998.

Automated Life Insurance Underwriting

The technology affiliate of the insurance company Lincoln Re is suing a software company, Allfinanz, for patent infringement. Allfinanz has a system that issues a temporary life insurance policy at a bank branch within 30 minutes. Lincoln Re alleges this system infringes two patents it has obtained for automated risk assessment and decision making. The company is seeking an injunction and damages.

Internet Banking

In 2000, S1 (formerly Security First Network Bank) obtained a patent on its three-tier financial transaction system. It is suing the company Corillion, a developer of Internet banking software used by several dozen large financial institutions. There are at least two dozen patents that involve online banking. Microsoft owns at least one of them.

ARE THESE NEW PATENTS GOOD OR BAD?

A patent system is a reward system developed to encourage inventors to invest the time and resources to make valuable discoveries. But, by creating temporary monopolies, patents also have social costs. In particular, consumers must pay more than the marginal cost of the new product or process, which results in too little use of the innovation (see *Patent Basics*). There are also transaction costs associated with determining the validity of patents and instances of infringement, plus the costs associated with negotiating licensing agreements. Has extending patent protection to computer software and business methods increased the rate of innovation? Is that increase worth the social costs associated with these patents?

Are the New Patents Stimulating Innovation? Between 1995 and 1998, spending on research and development (R&D) by firms in the computer programming and data processing industry increased 67 percent, reaching \$14.3 billion. The industry accounted for 10 percent of all private spending on R&D, one-third of all R&D spending by nonmanufacturing firms, in 1998. Employment of scientists and engineers by the industry, another measure of research activity, increased 59 percent, to 123,000. That was more than 12 percent of all scientists and engineers employed by industry (36 percent of those employed by nonmanufacturing firms) in 1998.¹⁶

As impressive as these numbers are, they largely reflect the rapid growth of the industry, whose sales rose 65 percent between 1995 and 1998.¹⁷ Another way to evaluate these trends is to examine the ratio of

¹⁶ The data on R&D spending and employment of scientists and engineers are from the National Science Foundation's annual survey of industrial R&D for firms contained in the standard industry classification (SIC) 737, which includes developers of custom-designed and prepackaged software, integrators of computer hardware and software, and firms that provide data processing services to other companies.

R&D spending to sales, which is a measure of the research intensity of an industry. A high ratio might imply there are potentially many new products or processes that are worth exploring.¹⁸ The ratio of R&D spending to sales for publicly traded companies in the software and data processing sector has increased from about 5 percent in the early 1980s to about 7.5 percent in recent years.¹⁹ Most of that increase occurred prior to the 1990s, a time when the patentability of computer programs was still uncertain. The percent increase in R&D intensity for this sector (59 percent) between 1980 and 1999 is comparable to the overall trend for publicly held companies (56 percent).

We have far less information about research activity in the financial services sector than in other parts of the economy. We do know that R&D spending and employment of scientists and engineers are tiny relative to total industry revenues and employment. But we also know that both have increased rapidly in recent years. Between 1995 and 1998, R&D spending in the financial services sector more than doubled, to \$1.6 billion, and employment of scientists and engineers more than tripled, to 17,500.²⁰

¹⁷ This calculation is based on the Census Bureau's Annual Survey of Service Industries.

¹⁸ This interpretation of the R&D/sales ratio can be criticized. Sales tend to change more rapidly than R&D budgets. If previous investments in R&D are very successful, sales growth will accelerate. That would depress the R&D/sales ratio until R&D budgets are revised upward. That is why it is important to examine this ratio over a number of years.

¹⁹ The R&D/sales ratio described here is derived from all publicly traded firms classified in SIC 737 in Standard and Poor's Compustat data set.

²⁰ Unlike firms in many other industries, most providers of financial services do not report their R&D spending, so the only available data of this sort come from the NSF survey. The NSF reports totals for firms contained in SIC 60-65 and SIC 67 (the finance, insurance, and real estate sectors). Unfortunately, that survey did not cover the service sector in detail prior to 1995.

It is important to remember that patents in these industries are a recent phenomenon. The patentability of computer software was established less than a decade ago. The patentability of business methods was only clearly established in 1998. Firms in the computer software and financial services industries were innovating rapidly long before it was thought possible to patent their innovations, yet they found effective ways to exploit their innovations without patents. This makes the availability of patent protection an unlikely explanation for the software revolution. It is even less likely that patent protection played an important role in the rapid financial innovation seen over the last three decades.

More Patents Are Not Necessarily Better. From the standpoint of theory, the claim that more patent protection encourages more R&D, and therefore more innovation, is a qualified one. In economic models where innovations build on each other over time, extending patent protection to less significant innovations can either raise or lower the rate of innovation, depending on characteristics of the industry.²¹

When patentability criteria are fairly strict, there is a significant risk that a discovery will not qualify for protection, which might discourage inventors from undertaking costly projects. If patentability requirements are relaxed, more inventions will qualify for patent protection, possibly reducing the risk of imitation that inventors face. But at the same time, competition between related patented technologies will increase, reducing the profits that patents generate. That would make patents less valuable and possibly reduce the incentive to innovate. In high technology industries, where innovation is already rapid, the first effect is weaker and the second effect is stronger. As a result,

²¹ The ambiguous effect of weaker patentability criteria is found in a number of theoretical models, including those developed by Jim Bessen and Eric Maskin, Ted O'Donoghue, and Olivier Cadot and Steve Lippman.

relaxing patentability criteria is more likely to reduce research activity in industries that are already innovating rapidly.²²

Declining Patent Quality.

Because patents have social costs, they should be granted only for inventions that are really new. Moreover, the property rights conferred by a patent should be reasonably related to what an inventor has actually discovered. It is the job of the patent office to ensure the inventor's claims do not overreach.

It is only natural that the patent office will make more mistakes when evaluating patent applications in a new technology field than in fields it is already familiar with. The patent examiner will not know where to look for the prior art, which will not be found in the records of previously issued patents, and will have more difficulty determining whether an applicant's specification is truly novel or nonobvious.²³ All of these problems appear in the case of computer programs.

The patent office has been widely criticized for issuing patents on garden-variety software technologies. The undisputed black eye occurred in 1993, when it granted a patent to Compton's Encyclopedia for a multimedia search and retrieval system. Unfortunately, the patent office was unaware of a great deal of prior art, including prior patents. The commissioner soon ordered the patent to be reexamined, and all its claims were eventually rejected.²⁴ Unfortunately, the patent office is apparently reliving

the experience. In 1998, it issued a patent for a method to eliminate problems associated with the year 2000 date change in computer programs. Again, there was an outcry that the patent covered techniques widely used for many years. And, again, the commissioner announced that the patent would be reexamined.²⁵

moving industry, a few months' delay can doom even a state-of-the-art product.

In this environment, firms may find themselves in a "patent arms race." This has both defensive and offensive qualities. On the one hand, firms believe they must patent as much as possible to prevent rivals from

The patent office has been widely criticized for issuing patents on garden-variety software technologies.

Some argue that the worst patents do not matter because no one would ever try to enforce them. But the real concern is that these examples signal a large quantity of important, but poorly examined patents that could lead to increased litigation, which is very costly.²⁶ But for every critic of these patents, there are others who justify them as a means of protecting small start-up firms or a new source of revenues for established ones.

A Patent Arms Race? The nature of innovation and recent patenting activity may also increase the potential costs associated with poor patent quality. In the computer software industry and many other high technology industries, innovation tends to be cumulative — that is, inventions tend to build on one another. As more and more of these inventions are patented, firms are finding they need to cross-license technologies, often from rivals. In the background there is always the fear that a preliminary injunction will delay a product's introduction. In a fast-

obtaining patents that might threaten commercialization of their inventions. Even when this strategy fails, the firm is likely to have a larger stock of proprietary technologies to trade in cross-licensing agreements with their rivals.²⁷ On the other hand, a well-constructed "patent thicket" can be used to raise barriers to entry, particularly for start-up firms.

At nearly the same time that computer-related inventions became patentable, the federal courts raised the presumption that the decisions of the patent office were correct. Litigants must now produce more evidence to overcome this presumption, which increases the cost of invalidating erroneous patents. In addition, the courts are more willing to grant preliminary injunctions than in the past, which increases the risk that a firm can be erroneously locked out of the market for a new good or service.

²² This somewhat counter-intuitive result is demonstrated in my 1999 Working Paper and explained in more detail in my 1999 *Business Review* article.

²³ At this point in the application process, the burden of proving that the claimed invention is obvious lies with the examiner. If the applicant is aware of any relevant prior art, he or she is obligated to disclose it in the patent application or risk having the patent invalidated. But applicants are not obliged to conduct a thorough search of the prior art before applying for a patent.

²⁴ Patent no. 5,241,671. See the article by E. Robert Yoches.

²⁵ The patent in question is no. 5,806,063 (Date Formatting and Sorting for Dates Spanning the Turn of the Century), granted to Bruce Dickens. The reexamination was ordered in December 1999.

²⁶ John Barton cites a recent American Intellectual Property Association survey, which found that litigating a typical patent infringement suit through the trial stage generated \$3 million in legal costs.

²⁷ In the mid 1990s, Wes Cohen and his colleagues asked over 1000 U.S. manufacturing companies why they patented their inventions. The following answers are ordered by how frequently they were cited: to prevent copying (96 percent); to prevent rivals from obtaining blocking patents (82 percent); to prevent infringement suits (59 percent); and to use in negotiations with other firms (47 percent). The percentages in parentheses are for product innovations. The order is the same for process innovations, but the percentages are lower. Bronwyn Hall and Rosemarie Ham Ziedonis also found that these factors were important in explaining the surge in patenting by American semiconductor firms.

This may be a prescription for more litigation, and it may deter smaller firms from entering certain markets, thereby reducing competition.

But It's Not Clear-Cut.

There are a few countervailing arguments to consider. First, while we should be concerned about the possibility of technological bottlenecks, we cannot be certain they will happen. Firms may develop more efficient ways to cross-license their proprietary technologies. For example, the copyright collectives ASCAP and BMI coordinate the collection and distribution of royalties for music played on the radio. A number of web sites that would facilitate patent licensing are already in development. Another approach is the formation of institutions that pool the intellectual property rights of their members. This is not a new idea—patent pools have been used to resolve technology bottlenecks in a number of industries. But any attempt to construct a patent pool will come under scrutiny by the Justice Department, which will be concerned about the possibility of collusion.²⁸

Nor do we know how these patents will hold up in the courts. The arguments given above suggest that patent holders are in a strong position. But defendants have a strong incentive to find compelling examples of prior art missed by the patent office. Patents of really poor quality are unlikely to stand up in court. Anticipating this, owners of such patents may not be very aggressive in asserting their rights.²⁹ Another possibility is that courts will narrow what they see as overly broad claims contained in

²⁸ Robert Merges' 1996 article describes how these arrangements work to reduce the transaction costs associated with sharing copyrighted content or patented technologies. But he also points out that ASCAP has operated under an antitrust consent decree since the 1950s.

²⁹ But there remains a legitimate concern that some patent owners will have an incentive to file nuisance suits if they believe that defendants will find it cheaper to settle rather than pursuing an expensive victory in court.

these patents. Rivals may find that it is easy to "invent around" a narrow patent on a computer program or a business model.³⁰

WHAT CAN BE DONE?

Barring significant new legislation or a sudden reversal of course by the courts, patents on computer software and business methods are here to stay. What, then, can we do to maximize the benefits and minimize the costs? It turns out there are many things we can do, but many of the following ideas are controversial.

More Resources. The obvious thing is to make sure the patent office has the resources and expertise required to do quality examinations in these new fields. The patent activities of the office are largely funded through fees charged to process patents and to keep them in force. The rapid increase in patent applications has contributed to a rapid rise in fees, now approaching \$1 billion a year. Even with this increase in resources, the patent office spends only about \$2700 per patent application processed. The amount of time required to process patents has increased more than 50 percent since 1994 and the backlog of pending patent applications has more than

³⁰ For example, just before Christmas 1999, Amazon.com successfully sued to prevent Barnes and Noble from using a one-click ordering system on its web site. While this was an embarrassing defeat in court, Barnes and Noble simply added a second mouse click to its program. Unisys holds a patent on a data compression technique integral to the GIF format often used for pictures on web pages. Its attempts to secure licensing revenues from this patent stimulated the adoption of new compression formats.

³¹ The number of applications awaiting action by an examiner increased from 107,000 in 1994 to 243,000 in 1999. In 1994, the average patent application took 16 months to process. In 1999, the average processing time was 25 months. The delay is worse in certain technology fields. It took an average of 31 months to issue a patent on communications or information processing technology compared to an average 26 months for all technologies. It should be noted that, throughout the 1990s, a share of the fees the

doubled.³¹ In certain technology areas, the increase in the number of examiners has not kept up with the increase in applications.

More Information. Many observers have called for the development of databases containing examples of nonpatent prior art that could be made available to patent examiners. The patent office is currently moving away from its paper-based system to an approach emphasizing computerized searches of publicly available databases. But the agency has experienced many technical problems in implementing its new search systems. The patent office and others have called upon the software industry to develop its own database of prior art. Indeed, the Software Patent Institute was founded for this very purpose in 1994, but apparently more needs to be done.

Other ideas are more controversial. For example, suppose we require that before filing for a patent, applicants must conduct a search of the prior art and report what they find in their application to the patent office.³² This would shift some of the burden of discovery back to the inventor. The problem is how such a requirement might be implemented. How do we set a standard for a minimally acceptable search of the prior art? How would foreign applicants comply? Do we want the patent office to be swamped with printouts from 250,000 web searches?

Currently, most of the burden of identifying relevant prior art lies with the patent office. Some scholars propose adopting an opposition system – an administrative process whereby a third party can dispute a patent either just before or just after it is issued.³³ The idea is to use the self-interest of

patent office collects has been diverted to the Treasury. That share has grown significantly in recent years.

³² Currently, it is not uncommon for applicants to search the prior art, but the motivation is self-interest rather than any binding legal obligation.

³³ See Robert Merges' 1999 article.

customers or competitors to generate more information than the patent office is able or willing to produce, which could improve the quality of patents. Opposition systems have been used for many years in Europe and, until recently, Japan. But they are not universally supported because the evidence presented may be biased and third parties may have an excessive interest in contesting patents, even good ones.

Change the Standards Used by the Courts? Recall that during the 1980s, the courts began to require clear and convincing evidence that a patented invention did not meet the statutory requirements before invalidating the patent. Many legal practitioners and scholars supported a more stringent burden of proof, arguing that the courts had been all too willing to second guess the patent office during the 1970s.

But today, it is not unreasonable to ask whether, for patents in new technology fields, the courts should be more skeptical about the quality of patent examinations. In that case, we may wish to return to the previous evidentiary standard (a preponderance of the evidence) when evaluating patents on technologies unfamiliar to the patent office. It would also seem prudent for courts to be more circumspect about granting preliminary injunctions before reaching a final conclusion about a patent's validity.

Stricter Patentability Criteria Would Help. One way to reduce the effect of issuing more erroneous patents is to adopt a more stringent test for nonobviousness. Under a more rigorous test, the fact that certain prior art was missed is less likely to affect the examiner's final decision. Nor is it at all clear that adopting more rigorous patentability criteria would adversely affect the incentive to innovate. As discussed in the previous section, there is little evidence that moving to weaker patentability criteria in the 1980s led to more innovation.³⁴

³⁴ See Adam Jaffe's review paper.

Small Changes Are Already Being Made. Recognizing that certain problems exist, the patent office is not standing still. It is developing special rules for examining patents in the areas of computer software and business methods. It is hiring new examiners trained in these fields and developing contacts with a variety of trade associations. It has also re-requested additional resources.


In 1999, Congress enacted a special prior use defense for patents on "methods of doing or conducting business." Under this defense, a firm that was using a business method, but had not disclosed it (because it was a trade secret), cannot be found to infringe the patent.³⁵ The usefulness of this exception is unclear, in part because the definition of "business method" must be established through litigation. Nor is it clear that encouraging firms to conceal innovative business practices or reducing the incentive of firms to dispute invalid patents would be socially beneficial.³⁶

³⁵ Section 4302 of the American Inventors Protection Act (Public Law 106-113), enacted in 1999. A patent cannot be invalidated by the existence of an earlier instance of the invention when it is concealed, as it must be protected as a trade secret.

³⁶ James Barney examines these and other issues in his recent article. One unanswered question is whether the prior use exception would apply in cases alleging infringement of a patent on computer software that implements a business method.

CONCLUSION

The traditional rationale for granting patents is that they are a reward to inventors and, as such, spur innovation. But this does not mean it is always better to have more patents. We don't know whether extending patent protection to computer programs and business methods implemented via computers will stimulate innovation in the software industry or, for that matter, the development of financial services on the Internet. But there are good reasons to expect that such patenting will not provide a whole lot more incentive to innovate than these firms already have. And it may well be the case that the costs associated with enforcing, licensing, or invalidating these patents could be higher than we have seen in other industries and in other eras.

We can do a number of things to minimize the problems associated with these patents and to maximize their benefits. We can at least take steps to improve the quality of patents being issued. This may involve additional resources, but it may also require a structural change in our patent process. And we should do more careful empirical research on the effects of increasing the availability of patents in high technology industries. This would give policymakers more and better information about the costs and benefits of the ongoing changes in our patent system. 

Cases Cited

Gottschalk v. Benson, 409 U.S. 63 (1972).

Diamond v. Diehr, 450 U.S. 175 (1981).

Hybritech Inc. v. Monoclonal Antibodies Inc., 802 F.2d 1367 (Fed. Cir. 1986).

In re Alappat, 33 F.3d 1526 (Fed. Cir. 1994).

Medtronics Inc. v. Intermedics, Inc., 799 F.2d 734 (Fed. Cir. 1986).

Paine, Webber, Jackson & Curtis, Inc. v. Merrill Lynch, Pierce, Fenner, and Smith, Inc., 564 F. Supp. 1358 (D. Del. 1983).

State Street Bank and Trust Co., Inc. v. Signature Financial Group, Inc., 149 F.3d 1368 (Fed. Cir. 1998).

REFERENCES

- Assistant Secretary of Commerce and Commissioner of Patents and Trademarks. Annual Report. Washington: U.S. Patent and Trademark Office, various years.
- Barney, James R. "The Prior Use Defense: A Reprieve for Trade Secret Owners or a Disaster for the Patent Law?" *Journal of the Patent and Trademark Office Society*, Vol. 82 (2000), pp. 261-73.
- Bessen, James, and Eric Maskin. "Sequential Innovation, Patents, and Imitation," MIT Economics Department Working Paper No. 00-01 (2000).
- Cadot, Olivier, and Steven A. Lippman. "Fighting Imitation with Fast-Paced Innovation," Working Paper 97/73, INSEAD (1997).
- 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)," NBER Working Paper 7552 (2000).
- Coolley, Ronald B. "The Status of Obviousness and How to Assert It as a Defense," *Journal of the Patent and Trademark Office Society*, Vol. 76 (1994), pp. 625-44.
- Cunningham, M.A. "Preliminary Injunctive Relief in Patent Litigation," *IDEA*, Vol. 35 (1995), pp. 213-59.
- Del Gallo, III, Rinaldo. "Are Methods of Doing Business Finally Out of Business as a Statutory Rejection?" *IDEA*, Vol. 38 (1998), pp. 403-37.
- Dunner, Donald R., J. Michael Jakes, and Jeffrey D. Karceski. "A Statistical Look at the Federal Circuit's Patent Decisions; 1982-1994," *Federal Circuit Bar Journal*, Vol. 5 (1995), pp. 151-80.
- Friedman, David D., William M. Landes, and Richard A. Posner. "Some Economics of Trade Secret Law," *Journal of Economic Perspectives*, Vol. 5 (1991), pp. 61-72.
- Gleick, James. "Patently Absurd," *New York Times Magazine*, March 12, 2000.
- Hall, Bronwyn, and Rosemarie Ham Ziedonis. "The Patent Paradox Revisited: Determinants of Patenting in the US Semiconductor Industry, 1979-95," mimeo, Wharton School of Business, University of Pennsylvania (2000).
- Hansen, Evan. "Patent Demands May Spur Unisys Rivals in Graphics Market," CNET News.com, April 18, 2000 (<http://news.cnet.com/news/0-1005-200-1713278.html>).
- Hunt, Robert M. "Nonobviousness and the Incentive to Innovate: An Economic Analysis of Intellectual Property Reform," Working Paper 99-3, Federal Reserve Bank of Philadelphia (March 1999).
- Hunt, Robert M. "Patent Reform: A Mixed Blessing for the U.S. Economy?" Federal Reserve Bank of Philadelphia *Business Review*, November/December 1999.
- Hunt, Robert M. "The Value of R&D in the U.S. Semiconductor Industry: What Happened in the 1980s?" mimeo, Congressional Budget Office, 1996.
- Jaffe, Adam B. "The U.S. Patent System in Transition: Policy Innovation and the Innovation Process," NBER Working Paper 7280 (1999).
- Kastriner, Lawrence G. "The Revival of Confidence in the Patent System," *Journal of the Patent and Trademark Office Society*, 73 (1991), pp. 5-23.
- Lotvin, Mikhail, and Barry D. Rein. "Patentability of Software after AT&T v. Excel: The Rollercoaster Ride Is Almost Over," *The Journal of Proprietary Rights*, Vol. 11 (1999), pp. 10-16.
- Merges, Robert P. "As Many as Six Impossible Patents Before Breakfast: Property Rights for Business Concepts and Patent System Reform," *Berkeley Technology Law Journal*, Vol. 14 (1999), pp. 577-615.
- Merges, Robert P. "Contracting into Liability Rules: Intellectual Property Rights and Collective Rights Organizations," *California Law Review*, Vol. 84 (1996), pp. 1293-93.
- Mester, Loretta J. "The Changing Nature of the Payment System: Should New Players Mean New Rules?" Federal Reserve Bank of Philadelphia *Business Review*, March/April 2000.
- National Science Foundation. *Research and Development in Industry: 1998*. (Arlington, VA, 2000).
- Nichols, Kenneth. *Inventing Software: The Rise of Computer Related Patents*. London: Quorum Books, 1998.
- O'Donoghue, Ted. "A Patentability Requirement for Sequential Innovation," *RAND Journal of Economics*, Vol. 29 (1998), pp. 654-79.
- Oliner, Stephen D., and Daniel E. Sichel. "The Resurgence of Growth in the Late 1990s: Is Information Technology the Story?" FEDS Discussion Paper 2000-20 (2000).
- Rai, Arti K. "Addressing the Patent Gold Rush: The Role of Deference to PTO Patent Denials," *Washington University Journal of Law and Policy*, Vol. 2, 2000, pp. 199-228.
- Thomas, John R. "The Patenting of the Liberal Professions," *Boston College Law Review*, Vol. 40 (1999), pp. 1139-85.
- U.S. Census Bureau. *Services Annual Survey: 1998*. Washington: U.S. Government Printing Office, 1999.
- U.S. Patent and Trademark Office. *Automated Financial or Management Data Processing Methods (Business Methods)*, A White Paper (July 19, 2000). (<http://www.uspto.gov/web/menu/busmethp/index.html>).
- U.S. Patent and Trademark Office. *Data Processing: Financial, Business Practice, Management, or Cost/Price Determination (Class 705) 1/1977-12/1999*. Technology Profile Report (January 2000).
- U.S. Patent and Trademark Office. *Internet-Related Patents 1/1977 - 6/2000*. Technology Profile Report (July 2000).
- U.S. Patent and Trademark Office. *Patent Counts by Class by Year January 1, 1977 - December 31, 1999* (April 2000).
- Yoches, E. Robert. "The Compton's Reexamination - A Sign of the Times," *The Computer Lawyer*, Vol. 12, No. 3 (March 1995), pp. 14-18.

Who Cares About Volatility?

A Tale of Two Exchange-Rate Systems

BY SYLVAIN LEDUC

Currency debacles in Mexico in 1994-95 and in Asia in 1997-98 suggest that exchange-rate policies may be very important for these economies. Some economists go so far as to argue that, for many small countries, the choice of an exchange-rate system may be their single most important macroeconomic policy decision.

The post-World War II experience of industrial countries, however, paints a different picture of the importance of exchange-rate policies for an economy's performance. In fact, the exchange-rate system mattered surprisingly little for the performance of these economies.

The debate over the benefits of different exchange-rate systems goes back a long way. In the 1940s, some economists argued that exchange rates determined by market forces would be very unstable — they would experience wild and erratic movements driven mostly by the speculative motives of investors. Critics of such flexible exchange-rate systems feared that

the uncertainty created by these movements in exchange rates would lead to a fall in trade between nations and, thus, to lower standards of living. Nobel laureate Milton Friedman disagreed forcefully with that view. In an important essay, which influenced the decision to adopt a system of flexible exchange rates in the early 1970s, Friedman argued that as long as underlying economic conditions are stable, there is no presumption that exchange rates will move excessively.

Although much has been written on the subject, the debate is far from settled. Indeed, the sheer number of different exchange-rate systems currently in place in the world demonstrates that policymakers disagree on the merits of different exchange-rate arrangements. For instance, as of June 1999, the International Monetary Fund reported that 67 countries pegged their currency, eight adhered to a currency board arrangement, 37 either used the currency of another country as the sole legal tender or belonged to a monetary union,

and the remaining 73 followed more flexible arrangements, such as managed or independent floating (see *Different Types of Exchange-Rate Regimes* for a short description of the different systems). Even among the more homogeneous group of 29 countries that make up the Organization for Economic Cooperation and Development (OECD), six were pegging their currencies, 12 followed arrangements of independent or managed floating, and 11 had just formed a monetary union in which they adopted a common currency, the euro, which will be the sole legal tender by 2002.

Exchange-rate arrangements vary across time as well as across countries. For instance, at the start of the 20th century, most Western countries' currencies were rigidly fixed to gold, under the system known as the gold standard, which collapsed with the outbreak of World War I. In the 1930s, many countries, facing the hardships of the Great Depression, opted for more flexibility and decided to let market forces determine the value of their currencies.¹ Then, in 1945, the architects of the Bretton Woods system struck a compromise by creating a system of fixed, but adjustable exchange rates, in an attempt to combine the benefits of the gold standard with those of flexible exchange rates (see *The Bretton Woods Exchange-Rate System*).

Countries spend much time and effort devising exchange-rate arrangements. One reason is that policymakers hope that the right exchange-rate system will help stabilize their economies. A stable economic environment decreases the amount of

¹ For a thorough account of the gold standard and the interwar period, see the book by Barry Eichengreen.



Sylvain Leduc is an economist in the Research Department of the Philadelphia Fed.

uncertainty people face when they have to make economic decisions such as how much to plant this season, how much to expand a factory, or how much to save for retirement. Is one particular exchange-rate system associated with a more stable economic environment? To shed some light on this question, we will review what happened historically to the volatility of certain macroeconomic variables for some industrial countries under different exchange-rate mechanisms. We will focus our attention on the Bretton Woods system of fixed, but adjustable exchange rates, in place from 1945 to 1971, and on the period since 1973, during which most industrial countries opted for flexible exchange rates.² We will see that, as the critics argued would happen, flexible exchange rates have been extremely volatile, but predictions of lower trade volumes, lower output, and lower standards of living failed to come true. Indeed, the volatility of exchange rates affected economies surprisingly little, whether we look at output, net exports, consumption, or investment.

WHAT ARE THE HISTORICAL FACTS?

Let's begin our investigation of how the exchange-rate mechanism affects macroeconomic variables by looking at some plots describing how those variables moved over time from 1957 to 1999.³ One important variable is the real exchange rate. People usually think about nominal exchange rates, which denote the price of one currency in terms of another. For instance, in January 1999 one U.S. dollar was worth 113 Japanese yen, but in January 2000, the U.S. dollar traded for 105 Japanese yen. Therefore, the dollar lost 7.1 percent of its value against the yen over that year, that is, the *nominal* exchange rate fell 7.1 per-

² This article does not study the period between August 1971 and March 1973 when the major industrial countries' currencies evolved under the Smithsonian Agreement.

³ The sample studied starts in 1957 instead of 1945 because many economic series are unavailable before that date.

cent. The *real* exchange rate, on the other hand, represents the relative prices of goods in two countries. A common measure used to calculate the real exchange rate is the nominal exchange rate multiplied by the ratio of consumer price indexes in the two countries. From January 1999 to January 2000, the consumer price index rose 3 percent more in the United States than in Japan, so the real exchange rate declined only 4.1 percent.⁴ In other words, while \$1.00 could buy 7.1 percent fewer yen in January 2000 than in January 1999,

⁴ In January 1999, the consumer price index in the United States was 164.7, and the consumer price index in Japan was 102, implying a real exchange rate of 182.5: the nominal exchange rate of 113 yen per U.S. dollar times the ratio of U.S. to Japanese price indexes. By January 2000, however, the U.S. consumer price index had risen to 169.1 while that in Japan had fallen to 101.5, resulting in a fall of the real exchange rate to 174.9.

\$1.00 of U.S. goods could be traded for 4.1 percent fewer Japanese goods in January 2000 than a year before.

The real exchange rate's volatility measures the extent to which the relative price of two countries' goods fluctuates over time. Since most countries switched from fixed to flexible nominal exchange rates in the early 1970s, real exchange rates have become much more volatile, as critics of flexible exchange rates predicted (Figure 1).⁵ However, the volatility of many other economic variables has not changed as much (Figure 2).⁶ If an

⁵ All variables have been detrended by taking the first difference of their logarithms; that is, we look at growth rates. For purposes of exposition, in Figures 1 and 2 the period between August 1971 and March 1973 is included with the flexible exchange-rate period.

⁶ See the article by Marianne Baxter and Alan Stockman for a more exhaustive study.

Different Types of Exchange-Rate Regimes

E

Exchange-Rate Arrangement with No Separate Legal Tender: Under this arrangement, two possible cases emerge. First, the currency of another country is used as the sole legal tender; for example, Panama uses the U.S. dollar. Second, the country may belong to a monetary union in which members of the union share the same legal tender; for example, the European countries that belong to the European Monetary Union share a common currency, the euro.

Currency Board Arrangement: In this case, the country is bound by law to exchange domestic currency for a particular foreign currency at a fixed exchange rate. Argentina uses this arrangement, exchanging one peso for one U.S. dollar.

Other Conventional Fixed Peg Arrangements: The country pegs its currency to that of another country or to a basket of currencies. Typically, the arrangement allows the exchange rate to fluctuate within a narrow band around a central rate.

Crawling Peg: As in the previous case, the country pegs its currency to that of another country, but it revises the exchange rate periodically at a fixed, pre-announced rate. Costa Rica uses a crawling peg system.

Managed Floating: The monetary authority of the country intervenes actively in the foreign-exchange market to influence the movements of the exchange rate. The monetary authority, however, does not specify or pre-commit to any particular value for the exchange rate.

Independent Floating: For the most part, the market determines the exchange rate of a country. The monetary authority rarely, if ever, intervenes in the foreign exchange market. England has a freely floating exchange rate.

The Bretton Woods Exchange-Rate System

I

n 1944, delegates from 44 countries met at Bretton Woods, New Hampshire, to reform the international monetary system.^a The delegates wanted to design a system that would combine the benefits of both flexible and fixed exchange-rate systems. The result was a system of fixed, but adjustable, nominal exchange rates.

Under the system, the U.S. dollar was fixed in terms of gold (initially at \$35 per ounce), and the U.S. Treasury bought and sold gold to maintain this official price. In turn, every other member country was to fix its currency to the dollar (and indirectly to gold) and keep its exchange rate within a 1 percent range on either side of the parity by buying or selling U.S. dollars in the foreign-exchange market. Only in the face of a significant and long-lasting deficit or surplus in its balance of payments was a country allowed to adjust the parity of its currency. Thus, the goal was to enjoy the stability associated with fixed exchange rates while simultaneously retaining the ability to move the nominal exchange rate when necessary to restore equilibrium in the balance of payments.

The system essentially collapsed in August 1971, when the U.S. suspended its promise to exchange gold for dollars at the official price.^b Many elements contributed to the fall of Bretton Woods, but an important one concerned the liquidity of the system. Under the agreement, the U.S.

Treasury fixed the price of the U.S. dollar in terms of gold by buying and selling gold on the market. In other words, the U.S. promised to exchange U.S. dollars for gold at the official price of \$35 per ounce. The system collapsed when other countries no longer believed that the U.S. could keep its promise to exchange U.S. dollars for gold at the official price. In the 1960s, U.S. reserves of gold steadily declined while the amount of U.S. liabilities to foreigners increased. That is, there were more and more U.S. dollars in circulation for every ounce of gold, putting more strain on the capacity of the United States to honor the agreement. Other countries, which had accumulated U.S. dollars, became afraid that the dollar would be devalued in terms of gold, and they started to convert their holdings of dollars into gold. In August 1971, President Richard Nixon suspended the convertibility of dollars into gold, which essentially ended the Bretton Woods system.

^a A collection of articles on the workings of the Bretton Woods system can be found in the book by Michael Bordo and Barry Eichengreen.

^b In 1971, however, the industrial countries were not yet ready to implement a system of flexible exchange rates. Under the Smithsonian Agreement, signed in December 1971, they adopted a system similar in spirit to that of Bretton Woods, although it allowed the exchange rates to fluctuate more. That system collapsed in March 1973.

observer didn't know the date on which the Bretton Woods system fell, it would be hard to tell from plots of industrial production when these countries switched to a flexible exchange-rate system. The same is true of other macroeconomic variables, including consumption, investment, or net exports.

Although a picture may be worth a thousand words, looking solely at figures like these may be deceiving. Table 1 reports the volatility of the real exchange rate between three countries and the United States, as well as, for each country, the volatility of its industrial production, consumption, investment, and net exports (all in inflation-adjusted terms) in the flexible exchange-rate period relative to their volatility in the Bretton Woods period.⁷ The table demonstrates that except for the real exchange rate, these economic variables are about equally volatile under the two different ex-

change-rate regimes. Certainly, since 1973 no variable has experienced an increase in its volatility similar to that of real exchange rates. Moreover, the increase in exchange-rate volatility did not lead to a fall in international trade and to lower standards of living, as critics of flexible exchange rates feared. In fact, the relationship between exchange-rate regimes and economic growth does not appear to be strong.⁸

SHOULD WE CARE ABOUT THE REAL EXCHANGE RATE?

The increase in the volatility of real exchange rates since 1973,

⁷ The results are not affected if we look at the variability of real exchange rates vis-à-vis a currency other than the U.S. dollar.

⁸ See the article by Atish Gosh, Anne-Marie Gulde, Jonathan Ostry, and Holger Wolf.

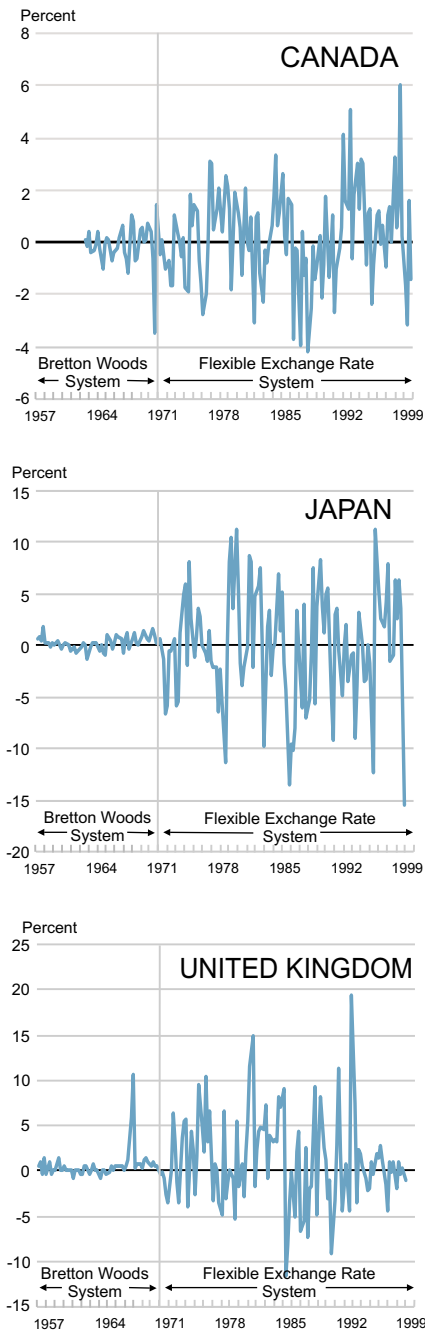
without similar increases in the volatility of other economic variables and with no adverse effect on trade volumes or standards of living, constitutes an important puzzle in international economics. The reason is that economists generally think that relative prices, like the real exchange rate, matter for allocating scarce resources efficiently.

Imagine, for a moment, that the U.S. economy does not trade with any other country and the relative price of cars suddenly increases. This increase would give people an incentive to switch expenditures from cars to other goods in the economy, therefore lowering production and employment in the car industry.

In theory, real exchange rates work just like the relative price of cars in the example above. The only difference is that the real exchange rate represents the relative price of goods between countries. Imagine now that the

FIGURE 1

Quarterly Growth Rates of Real Exchange Rates: Before & After Bretton Woods*

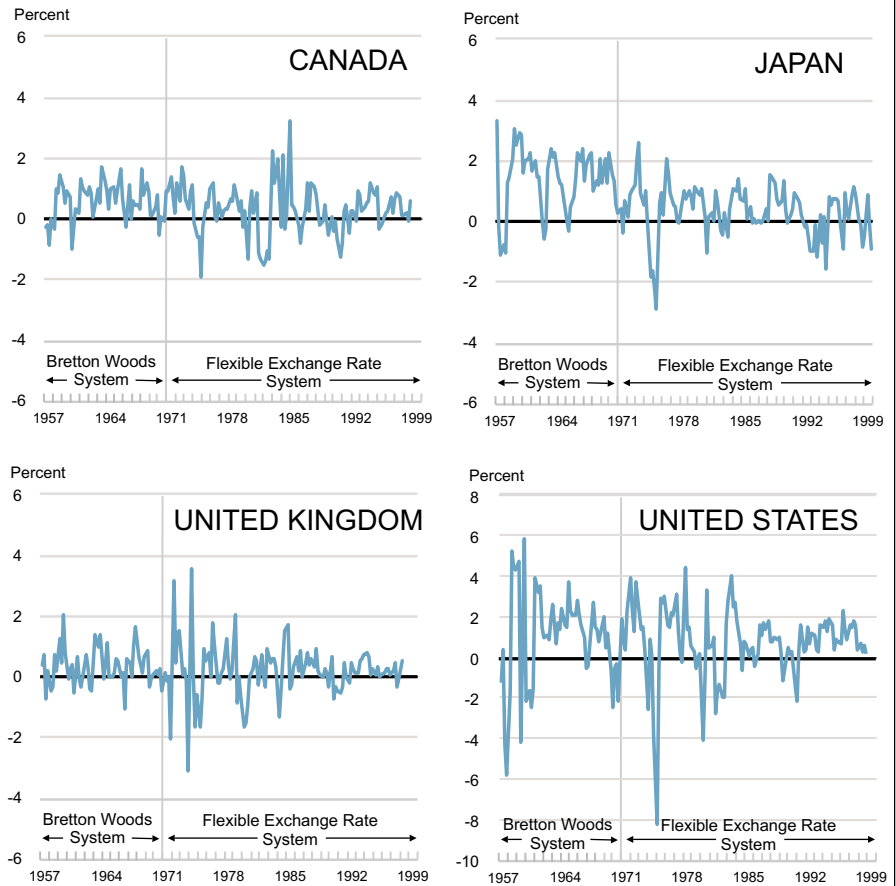


Source: IMF International Finance Statistics and Statistics Canada

* Real exchange rates shown vis-à-vis the United States

FIGURE 2

Quarterly Growth Rates Of Industrial Production: Before & After Bretton Woods



Source: IMF International Finance Statistics and DRI

U.S. economy opens up to trade, and suppose that the relative price of U.S. goods increases (i.e., the real exchange rate appreciates). Just like the increase in the relative price of cars in the previous example, the appreciation of the real exchange rate for U.S. goods should be associated with a shift in world demand away from U.S. products toward foreign-produced goods and, consequently, production and employment in the United States would fall.

These potentially large movements in production and employment

are what concern international economists and policymakers. According to this simple theory, movements in real exchange rates should coincide with movements in resources in the world economy. So if the volatility of the real exchange rate increases, the volatility of other economic variables, such as output or consumption, should also increase. However, as the discussion above showed, this clearly did not happen after 1973. Economists have been trying to solve this puzzle for some time now. Our search for a solution will be helped by learning whether the

TABLE 1: Ratio of Volatility

Flexible Exchange-Rate Period Relative to Bretton Woods Period

	Real Exchange Rate	Output	Consumption	Investment	Net Exports
Canada	3.42	1.53	1.03	1.18	0.79
Japan	8.84	0.81	1.04	0.74	0.70
United Kingdom	3.03	1.31	0.99	0.94	0.88
United States		0.71	0.95	1.02	0.92

Sources: IMF International Financial Statistics, DRI, and Statistics Canada
 The volatility of a variable is measured by the standard deviation of quarter-to-quarter growth rates of that variable. Quarter-to-quarter growth rates are calculated as the change from each quarter to the next in the logarithm of the variable.
 The Bretton Woods period covers 1957Q1 to 1970Q4, while the flexible exchange-rate period covers 1973Q1 to 1999Q4.

choice of exchange-rate systems affects the volatility of real exchange rates, or vice versa.

WHAT CAUSES WHAT?

Higher volatility of real exchange rates is associated with a flexible exchange-rate system. But what is cause and what is effect? Could it be that real exchange rates have become more variable since 1973 because the underlying circumstances affecting the economy have also become more variable, and consequently, countries decided to adopt a flexible exchange-rate system to insulate their economies from external shocks? Interestingly, the adoption of a flexible exchange-rate system in the early 1970s coincided with the first OPEC oil-price shock. Thus, real exchange rates may have become more volatile since 1973 because world economies have been subject to more real shocks, of which oil-price shocks are a prime example, and countries responded to these real shocks by moving to a flexible exchange-rate system. Perhaps the more volatile real exchange rate and the adoption of a flexible exchange-rate system resulted from an increase in the size of shocks to the economy. Although we do not have formal statisti-

cal results proving or disproving this case, the historical experience of Canada and Ireland shows that this is unlikely.

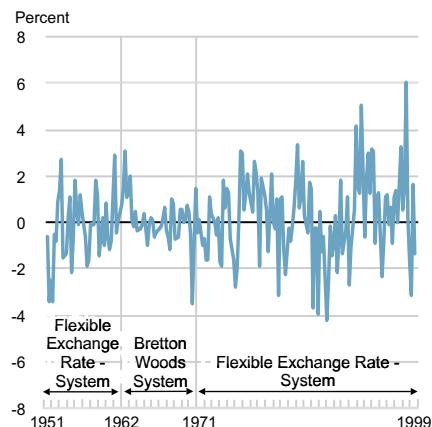
In 1950, Canada decided to leave the Bretton Woods system of fixed, but adjustable, exchange rates because it had difficulty setting a stable and credible exchange rate. The Canadian government had increased the value of the Canadian dollar, relative to other currencies, in 1946 and decreased it in 1949. In October 1950, facing strong market pressures toward an appreciation, Canada decided to let its currency float. The Canadian exchange rate was flexible until early 1962, when Canada rejoined the Bretton Woods system. Comparing the behavior of the Canadian exchange rate during these different periods shows that it is likely that the choice of exchange-rate system influences the behavior of a country's real exchange rate (Figure 3). The volatility of the Canadian real exchange rate was much lower when Canada was part of the Bretton Woods system in the 1960s than it has been under the current regime or than it was in the 1950s. In fact, each time the Canadian currency has been allowed to float, the real exchange rate has been roughly three

times more volatile than it was under the Bretton Woods system (Table 2).

Ireland's experience provides another example of the effects of the exchange-rate system on the economy. Until the end of 1978, the Irish pound was pegged to the British pound. But in January 1979, Ireland joined the European Monetary System (EMS), in which the Irish pound was effectively tied to the German mark. As in the Canadian case, the volatility of the real exchange rate between two countries is closely linked to the type of exchange-rate arrangement these countries follow (Table 3). Since the Irish pound was allowed to float against the British currency in 1979, the real exchange rate between Ireland and the UK has been more than twice as volatile as it was in the period from 1973 to 1978, when the currencies were tied to each other. The opposite pattern emerges between Ireland and Germany: after the Irish pound was essentially tied to the German mark in 1979, the volatility of the real exchange rate between Ireland and Germany fell by nearly one-half.

FIGURE 3

Quarterly Growth Rates of the Canadian Real Exchange Rate



Source: Statistics Canada and DRI

TABLE 2: The Canadian Experience

Standard Deviation of the Real Exchange Rate

	1950 - 1962	1962 - 1970	1971 - 1999
Real Exchange Rate	1.50	0.54	1.83

Source: Statistics Canada and DRI.

TABLE 3: The Irish Experience

Standard Deviation of the Real Exchange Rate

	1973- 1978	1979 - 1999
Irish Pound / UK Pound	1.69	4.21
Irish Pound / German Mark	4.87	2.77

Source: DRI

The numbers reported are the standard deviation of quarter-to-quarter growth rates of the real exchange rate. These quarter-to-quarter growth rates are calculated as the change from each quarter to the next in the logarithm of the real exchange rate.

From the experiences of Canada and Ireland, it's apparent that the change to a flexible exchange-rate system causes increased volatility in the real exchange rate, not vice versa.

WHY DO REAL EXCHANGE RATES EXPERIENCE WILD SWINGS?

One proposed explanation is that market psychology, not economic fundamentals like supply and demand, causes nominal and real exchange rates to move so much in a flexible exchange-rate system. Under this approach, the exchange rate becomes a self-fulfilling prophecy. For instance, suppose that a trader in the foreign-exchange market buys U.S. dollars because he expects the U.S. dollar to appreciate. If all traders have the same expectations and all decide to buy U.S. dollars, their actions push the value of the U.S. dollar up, confirming their

expectations. Thus, market psychology can lead to exchange-rate volatility.

But, in general, people do not like the uncertainty generated by more volatile exchange rates. Some use costly means, like hedging, to protect themselves against it.⁹ Since the rest of the economy behaves similarly un-

⁹ Hedging refers to the means investors and firms take to protect themselves from possible movements in currencies. For instance, suppose an American firm exporting to Canada expects to receive a payment of 100,000 Canadian dollars in three months. To protect itself from the possible depreciation of the Canadian dollar, the American firm could hedge by entering into a contract with a bank stipulating that the firm agrees to sell 100,000 Canadian dollars for U.S. dollars to the bank in three months, at a fixed rate of exchange set at the time the contract is agreed upon. Changes in the exchange rate between the Canadian dollar and the U.S. dollar during the three months will have no effect on the firm's profits.

der fixed or flexible exchange-rate regimes, fixing the nominal exchange rate could thus be beneficial in avoiding wild and irrational movements in exchange rates and their hedging costs. The market-psychology approach says that, left to itself, the market for foreign exchange does not work very well — exchange rates are too volatile, which imposes costs on the economy. Generally, however, economists believe in market forces and prefer explanations based on economic fundamentals rather than psychology. Commenting on the possibility that flexible exchange rates would be extremely unstable, Milton Friedman noted that “[the] advocacy of flexible exchange rates is not equivalent to advocacy of unstable exchange rates. The ultimate objective is a world in which exchange rates, while *free* to vary, are in fact highly stable. Instability of exchange rates is a symptom of instability in the underlying economic structure.”

Are the wild movements in real exchange rates since the early 1970s consistent with theoretical models based on economic fundamentals? The simplest model of exchange-rate determination, known as purchasing power parity, states that nominal exchange rates should move to offset inflation differentials across countries, leaving real exchange rates constant over time. This simple theory cannot explain the high volatility of real exchange rates.

What is happening is that the law of one price fails.¹⁰ The law of one price states that a good should sell for the same price in two different countries, when the prices are expressed in the same currency, after adjusting for tariffs and transport costs. If that were not the case, an individual would have an incentive to buy the good in the country where it is cheaper and sell it in the other country, an action called arbitrage in economic lingo. Such an individual is referred to as an arbitrageur.

¹⁰ See the articles by Charles Engel and by John Rogers and Michael Jenkins.

For instance, imagine that the same computer sells for 2000 Canadian dollars in Canada and for 1000 U.S. dollars in the United States. Moreover, suppose that 1 U.S. dollar can be exchanged for 1.5 Canadian dollars. Therefore, converting the U.S. price of the computer into Canadian dollars using the exchange rate, we find that 1500 Canadian dollars could buy the 1000 U.S. dollars needed to acquire this computer in the United States. To keep the example simple, let's assume there are no transport costs to ship a computer from the United States to Canada. An arbitrageur could make a profit of 500 Canadian dollars (333 U.S. dollars) by buying the computer in the United States and selling it in Canada. With sufficient arbitrage, prices, when expressed in the same currency, should converge, since arbitrageurs would raise the demand for computers in the United States and increase their supply in Canada.

If all sectors in the economy produce freely traded products and sell them in very competitive markets, the law of one price should hold, after adjusting for transport costs, and real exchange rates should not vary much. Rogers and Jenkins, however, found that 81 percent of the movements in real exchange rates are due to a failure of the law of one price. Moreover, Charles Engel and John Rogers showed that, for a wide range of commodities, the presence of transport costs cannot account for failure of the law of one price. This suggests that to better understand movements in real exchange rates, we need to have a better understanding of what causes the law of one price to fail.

THE SEARCH FOR A GOOD MODEL

Economists have been trying to develop a good model to explain the facts about real exchange rates. One important element of an explanation is that prices are slow to adjust to changes in the economy (often referred to as price stickiness). For instance, Figure 4 shows the nominal and the real exchange rates between Canada and the United States, as well

as the ratio of American to Canadian prices (as measured by the consumer price indexes). The high volatility of exchange rates since the early 1970s, combined with the nearly constant ratio of foreign to domestic prices, illustrates that the prices of goods are slow to adjust to changes in the economy, compared with financial variables such as nominal exchange rates.

Many researchers think that any model of exchange rates should include features that allow the prices

associated with highly volatile real exchange rates. While this may explain why real exchange rates are so volatile, it doesn't explain why more volatile real exchange rates aren't associated with more volatile consumption, output, and net exports. Therefore, the sluggishness of prices doesn't provide a full explanation of the puzzle.

It turns out that an explanation for the failure of the law of one price can help us understand why more volatile movements in the real exchange rate are not associated with more volatile economic fundamentals. This explanation, first postulated by Paul Krugman in a 1987 article, relies on what is known as pricing-to-market, which occurs when a firm sells the same product at different prices in different markets, as in our earlier example of computer prices. To determine if a firm set a different price in different markets, we would need to convert the price in one country into the other country's currency, using the exchange rate. A company might set different prices in different countries because of a difference in how strongly quantity demanded reacts to a change in price.

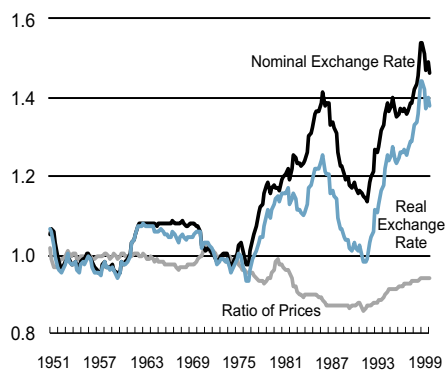
Of course, the extent to which a company can price to market may be limited by the possibility of arbitrage. With sufficient arbitrage, the difference in prices should vanish. Therefore, pricing-to-market is more likely to be found in industries, such as the car industry, in which products are tailored to local requirements.

Paul Froot and Paul

Klemperer provide another explanation for pricing-to-market. Their model relies on the observation that future demand for some firm's product may depend on the firm's current market share. In their framework, the higher the current market share, the higher the future demand will be. Such a relationship between current market share and future demand could arise because it is costly for consumers to switch brands. For example, a consumer may be unwilling to substitute a brand he has little information on for one he has tried and liked.

FIGURE 4

Canadian Exchange Rates and Price Ratio 1951-1999



Source: Statistics Canada and DRI

of goods to respond slowly to changes in the economy. Suppose some change in the economy causes a change in the nominal exchange rate. Since the prices of goods are slow to react (and since the real exchange rate equals the nominal exchange rate times the ratio of price indexes), real and nominal exchange rates should move approximately in line with each other. Therefore, flexible exchange-rate systems, in which nominal exchange rates are highly volatile, should also be

One property of pricing-to-market is that firms do not necessarily pass through movements in nominal exchange rates to the prices they charge in foreign countries, especially if the change in the exchange rate is temporary. Imagine, for example, that a U.S. firm sells its product in the United States and in Canada and that it is able to price to market. Suppose, moreover, that, as Froot and Klemperer argue, the firm's future demand for its product in each country depends on its current share of the market in each country. If the Canadian dollar were to temporarily depreciate vis-à-vis the U.S. dollar, the U.S. firm's profits would fall. The firm would get fewer U.S. dollars in exchange for each Canadian dollar it earns. The firm might react to the depreciation of the Canadian dollar by raising the price it charges Canadian consumers; that is, the U.S. firm might *pass through* the depreciation of the Canadian dollar to Canadian consumers. However, if the firm cares about its market share in Canada, it may prefer to keep constant the price it charges there, resulting in a cut in the firm's current profits.

In other words, pricing-to-market can engender price stickiness in local markets. Because of that stickiness, movements in the nominal exchange rate get translated into movements in the real exchange rate. Thus, pricing-to-market helps account for the high volatility of real exchange rates since 1973. And pricing-to-market can also help explain why the rest of the economy is unaffected by large movements in real exchange rates. Since, under pricing-to-market, movements in nominal exchange rates do not necessarily get passed through to consumer prices, consumers would have no incentives to switch from one good to another. Therefore, we would not expect production, consumption, investment, and net exports to respond strongly to exchange-rate movements if most firms price to market. Consequently, the high volatility of exchange rates would not be transmitted to other variables in the economy. In fact, many studies have uncovered pricing-

to-market in manufacturing industries. For instance, a study by Joseph Gagnon and Michael Knetter found that, instead of changing the price that they charge for their products in the United States, Japanese automobile exporters offset 70 percent of exchange-rate changes by adjusting profits.

Recently, researchers have built numerical models incorporating information on various aspects of the world economy to investigate how much changes in economic fundamentals (for instance, changes in monetary policy or movements in productivity) can account for the movements in exchange rates since 1973. The first wave of such models put the emphasis on perfectly competitive industries and movements in fundamentals driven primarily by changes in productivity across countries.¹¹ These frameworks were only partially successful at explaining the behavior of exchange rates, especially their high volatilities. In particular, they could not explain the large movements in real exchange rates that are due to the failure of the law of one price.

As a result, a second generation of numerical models tries to make sense of the large movements in

building on these previous studies, combined pricing-to-market and price stickiness to study the effects of exchange-rate regimes on the economy and found that these features could explain why the higher variability of real exchange rates since 1973 did not get transmitted to other economic variables. Figure 5, which shows the simulated time series for the real exchange rate, output, and net exports from our model, indicates that the model captures the empirical facts illustrated in Figures 1 and 2. Looking at the movements of the real exchange rate, an observer would be able to easily select the date at which the switch in exchange-rate regime occurred. However, looking only at the simulated series for output and net exports from our model economy, the observer would likely have a more difficult task.

CONCLUSION

Does the exchange-rate system matter? Looking at the Bretton Woods system and the flexible exchange-rate system that followed, this article showed that although real exchange rates have been much more volatile under the current flexible exchange-rate system, this high

Japanese automobile exporters offset 70 percent of exchange-rate changes by adjusting profits.

exchange rates under the current flexible exchange-rate regime by emphasizing imperfectly competitive industries with price stickiness and pricing-to-market, and changes in monetary policy.¹² These models have had relatively more success in explaining exchange-rate movements.

Recently, Luca Dedola and I,

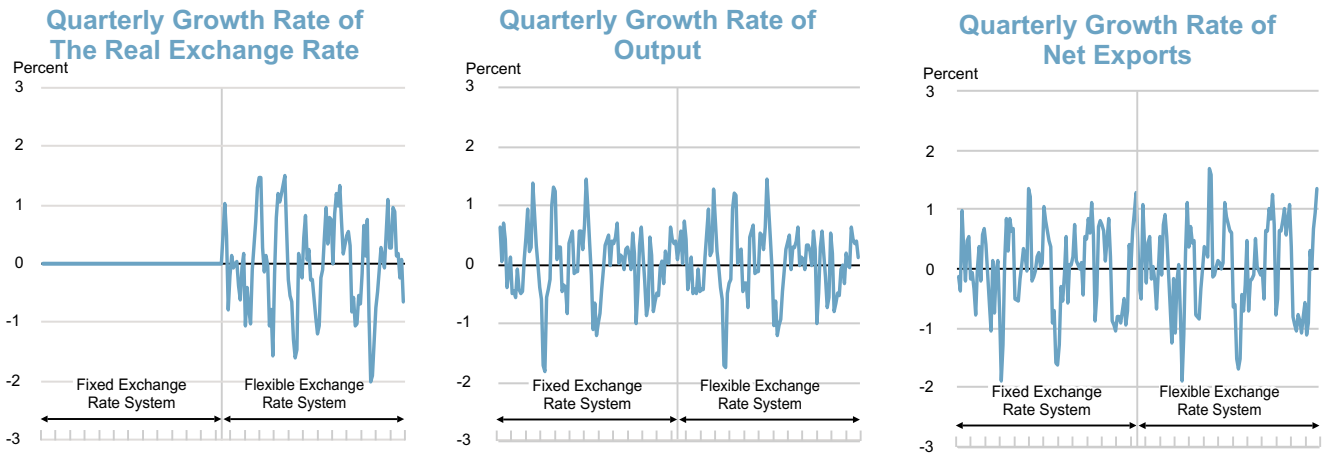
¹¹ See the article by David Backus, Patrick Kehoe, and Finn Kydland.

¹² See the articles by Caroline Betts and Michael Devereux; V.V. Chari, Patrick Kehoe, and Ellen McGrattan; and Robert Kollman.

volatility has not been transmitted to other sectors of the economy, at least not in the world's major industrial countries. In a sense, the early critics of flexible exchange-rate systems were half right. They were correct in predicting that exchange rates would be highly volatile if market forces determined them. However, their prediction of a lower trade volume, lower output, and overall lower standards of living did not materialize after the demise of the Bretton Woods system. In fact, in his 1989 book, Paul Krugman argues that flexible exchange rates "can move so much precisely because they seem to matter so little."

FIGURE 5


Model's Simulated Data Under Fixed & Flexible Exchange-Rate Systems*



* All the variables have been detrended

Does that mean that any type of exchange-rate system would suit any country? Probably not. Depending on the economic situation, a country could be harmed by its choice of an exchange-rate system. For instance, some researchers have shown that countries that left the gold standard in the early years of the Great Depression suffered much less than countries that kept their currency fixed to gold (see the book by Barry Eichengreen). This finding provides some evidence that the exchange-rate system matters, at least in drastic situations, and that the efforts of policymakers and academics to devise and understand different exchange-rate arrangements are important. Moreover, the recent experience of some emerging markets, such as East Asia or Latin America, suggests that exchange-rate volatility may very well matter for small, open economies, even though it does not

seem to matter much for larger, industrial countries. Indeed, a recent study by Shinji Takagi and Yushi Yoshida shows that Japanese firms exporting to East Asia, to a very large extent, do not price to market. As a result, movements in nominal exchange rates get transmitted into local

prices, which then affect consumption, production, and employment. Thus, unlike residents of industrial countries, those in small, open economies may very well care about exchange-rate volatility and which exchange-rate system is in place in their countries. 



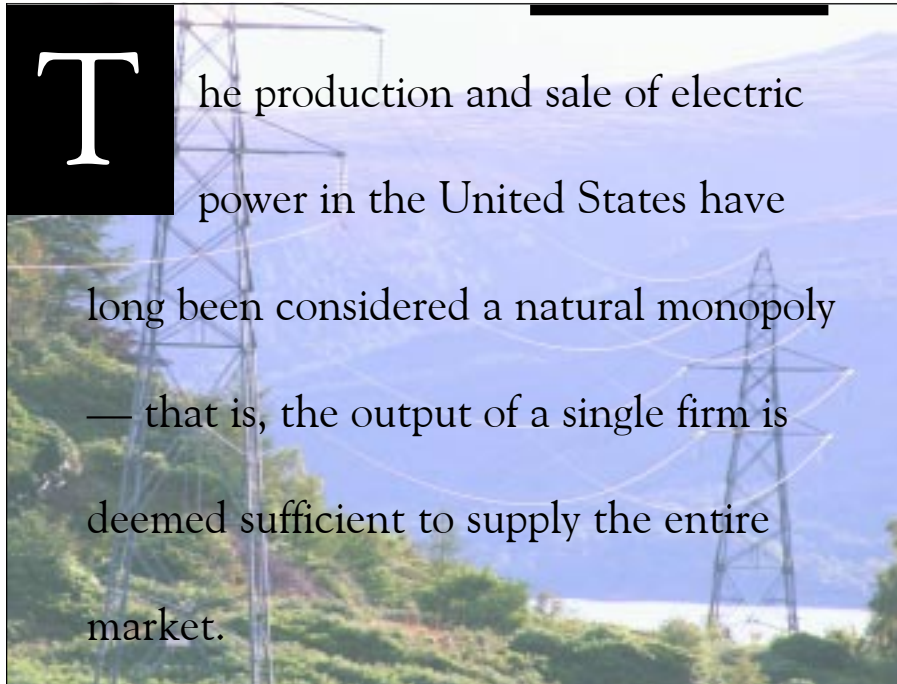
REFERENCES

- Backus, David, Patrick J. Kehoe, and Finn E. Kydland. "International Real Business Cycles," *Journal of Political Economy*, 100 (1992), pp. 745-75.
- Baxter, Marianne, and Alan C. Stockman. "Business Cycles and the Exchange-Rate Regime: Some International Evidence," *Journal of Monetary Economics*, May 1989, pp.377-400.
- Betts, Caroline, and Michael B. Devereux. "Exchange Rate Dynamics in a Model of Pricing-to-Market," *Journal of International Economics* (2000), pp. 215-44.
- Bordo, Michael D., and Barry Eichengreen. *A Retrospective on the Bretton Woods System: Lessons for International Monetary Reform*. Chicago: The University of Chicago Press, 1993.
- Chari, V. V., Patrick J. Kehoe, and Ellen R. McGrattan. "Monetary Shocks and Real Exchange Rates in Sticky Price Models of International Business Cycles," Federal Reserve Bank of Minneapolis Staff Report 223 (1998).
- Dedola, Luca, and Sylvain Leduc. "On Exchange-Rate Regimes, Exchange-Rate Fluctuations, and Fundamentals," Federal Reserve Bank of Philadelphia Working Paper 99-16.
- Eichengreen, Barry. *Golden Fetters: The Gold Standard and the Great Depression, 1919-1939*. London: Oxford University Press, 1995.
- Engel, Charles. "Is Real Exchange Rate Variability Caused by Relative Price Changes? An Empirical Investigation," *Journal of Monetary Economics* 32 (1993), pp. 35-50.
- Engel, Charles, and John H. Rogers. "How Wide Is the Border?" *American Economic Review* 86 (1996), pp. 1112-25.
- Friedman, Milton. *Essays in Positive Economics*. Chicago: The University of Chicago Press, 1953.
- Froot, Paul A., and Paul Klemperer. "Exchange Rate Pass-Through When Market Share Matters," *American Economic Review* 79 (1989), pp. 637-54.
- Gagnon, Joseph E., and Michael E. Knetter. "Markup Adjustment and Exchange Rate Fluctuations: Evidence From Panel Data on Automobile Exports," *Journal of International Money and Finance* 14 (1995), pp. 289-310.
- Gosh, Atish R., Ann Marie Gulde, Jonathan D. Ostry, and Holger Wolf. "Does the Exchange-Rate Regime Matter for Inflation and Growth?" IMF Economic Issues 2, International Monetary Fund, 1996.
- Kollman, Robert. "The Exchange Rate in a Dynamic-Optimizing Current Account Model with Nominal Rigidities: A Quantitative Investigation," IMF Working Paper 97/7 (1997).
- Krugman, Paul R. "Pricing to Market When the Exchange Rate Changes," in S.W. Andt and J.D. Richardson, eds., *Real-Financial Linkages Among Open Economies*. Cambridge, MA: MIT Press, 1987, pp. 49-70.
- Krugman, Paul R. *Exchange-Rate Instability*. Cambridge, MA: MIT Press, 1989.
- Rogers, John H., and Michael Jenkins. "Haircuts or Hysteresis? Sources of Movements in Real Exchange Rates," *Journal of International Economics* 38 (1994), pp. 339-60.
- Takagi, Shinji, and Yushi Yoshida. "Exchange Rate Movements and Tradable Goods Prices in East Asia: An Analysis Based on Japanese Customs Data, 1988-98," IMF Working Paper 99/31 (1999).

Rewiring the System:

The Changing Structure of the Electric Power Industry

BY TIMOTHY SCHILLER



The production and sale of electric power in the United States have long been considered a natural monopoly — that is, the output of a single firm is deemed sufficient to supply the entire market.

Monopoly electric utilities were permitted in order to achieve economies of scale, but they were subject to government regulation to prevent abuses of monopoly power. The primary regulators of electric utilities on the consumer side have been the public utility commissions in each state. Wholesale delivery of electric power from one utility to another has been under federal regulation. Now, the structure of the electric utility industry has begun to change in a direction that may ultimately lead to an open market national in scope.



Tim Schiller is an economic analyst in the Research Department of the Philadelphia Fed.

In the last few years significant changes have been made in federal and state laws that cover the industry. In 1998, Rhode Island and Massachusetts opened up their electric power markets to competition. The three states of the Third Federal Reserve District soon followed suit. Pennsylvania began a phased transition in 1998 that is now complete. In 1999, New Jersey opened its electric power market, and Delaware began to phase in competition among electric power suppliers. Since then, the restructuring of electric power markets has picked up momentum. As of October 2000, 23 states have enacted restructuring legislation. Prior to enactment of the new laws, businesses and households had no choice but to purchase electricity at regulated rates from the state-approved monopoly supplier of electricity. Now, in the restructuring states, purchasers of electric power may choose from among a number of state-approved producers or marketers that compete for customers.

Recent increases in fuel prices and stronger than expected demand for electricity across the country have led to increases in prices for electricity provided by nonregulated suppliers in some states, such as California, where power plant construction by monopoly utilities has lagged electricity demand. According to proponents of the new competitive structure, the current shortfall in supply will be overcome as more generating capacity is built and the electric power industry expands to become a truly national market. In the meantime, higher and more volatile prices are likely until new capacity catches up with demand.

Changes in the electric power industry have not come about overnight. Over the past 30 years, changes in energy markets and power technology and developments in economic theory have converged to produce a rethinking of the nature of electric utilities and a movement to revise the regulations that apply to them. This article reviews developments in public policy and technology that prompted the restructuring of the electric utility industry now under way, describes the new regulatory framework, and looks at what other developments may lie ahead.

THE IMPETUS FOR CHANGE

Why is the electric power industry being restructured now, after nearly a century of regulation as a natural monopoly? (See *Traditional Regulation of Electric Utilities* for a summary of the regulatory structure under which electric utilities operated until new laws were enacted.) To understand these developments, we must look at changes that have occurred in energy technology and markets during the past several decades as well as the evolution of

Traditional Regulation of Electric Utilities

T

he first electric utilities in the United States began operation in the 1880s as small generators and suppliers of electricity to city neighborhoods. Municipalities around the nation regulated entry into the industry through franchises. These franchises gave utilities the right, often for a specified number of years, to supply electricity within defined areas. In general, these rights were not exclusive. Thus, competing utilities would often serve identical or overlapping areas. As the technology of electric power generation and transmission developed, utilities began to serve larger areas, including entire cities. Concentration increased as utilities merged in some markets and as one or a few dominant utilities gained market share. By the first decade of the 20th century, increasing concentration prompted a movement toward state and federal regulation.

The first states to undertake regulation of electric utilities were New York and Wisconsin in 1907. State regulation had replaced municipal regulation in most states by 1914. Laws establishing state regulation were broadly based on the Wisconsin law, which established an independent state commission to regulate utilities. The commission controlled entry into the electric utility industry by requiring that new utilities obtain a certificate of convenience and necessity. In other words, the commission decided when and where new utilities could be established. The commission set service standards and rates, and it had authority over utilities' corporate structure and financial arrangements. The commissions reviewed utilities' operations and finances, inspected utilities' operations, and responded to complaints about service or safety from consumers. Either in response to utilities' requests, or on its own, the commission reviewed and proposed changes in electric rates.

Consolidation among utilities continued in the 1920s and 1930s, and corporations were formed that controlled utilities in several states. These corporations came under federal regulation with the passage of the Public Utility Holding Company Act (1935). This law gave the Securities and Exchange Commission detailed control over utilities' corporate structure. The commission has approval authority over holding companies' issuance of securities, ownership of assets, and dealings among subsidiaries. The law prohibited utilities from engaging in businesses not related to the production or transmission of electric power. The law also established the Federal Power Commission (renamed the Federal Energy Regulatory Commission in 1977) to regulate

utilities involved in interstate wholesale marketing or transmission of electric power (the sale or delivery of electric power from one utility to another). Another federal regulator of electric utilities is the Nuclear Regulatory Commission. Although primarily concerned with regulating the construction and operation of nuclear reactors, the commission also applies antitrust law when it considers a utility's application for a license for a nuclear reactor. *

The state and federal regulatory structure established in the early 20th century remained largely unchanged until the century was nearly over. Public policy, under both state and federal governments, was based on the theory of natural monopoly: electric power was most cheaply provided by a single supplier at all levels of production and distribution because of large fixed costs (capital investment in generation, transmission, and distribution facilities) and economies of scale. Public acceptance of this market outcome in order to achieve lower costs was accompanied by regulation intended to protect consumers from monopoly abuses in pricing.

Rather than simply dictate prices for electricity, state commissions tried to establish the requisite size of the sole supplier of electricity for the franchise areas (in terms of capital investment), then set rates to ensure the utility earned a market rate of return on its capital investment. However, over time, rates came to be set at various levels for various classes of users. State commissions defined types of users — such as residential, industrial (typically large manufacturing plants), and commercial (such as stores and office buildings) — and size classes and set rates at different levels for them, often favoring large users.

For most of their history, utilities remained vertically integrated, franchised, and regulated monopolies. There were occasional calls to reform the industry's structure, and some modifications were made to electric power pricing schemes, notably the introduction of peak-period prices in the late 1970s. Nevertheless, the federal and state regulatory structures put into place nearly a century ago prevailed until the pro-competition changes described in this article were put into effect.

* A brief summary of the state and federal regulatory structure may be found in Claire Holton Hammond, "An Overview of Electric Utility Regulation," in *Electric Power: Deregulation and the Public Interest*, John C. Moorhouse, editor, San Francisco: Pacific Research Institute for Public Policy, 1986, pp. 31-61.

economic thinking about industrial organization. An especially important development was the formation of regional transmission grids that reduced the need for every electric utility to have enough capacity to supply all the power needed in its

service area at times of exceptionally high usage. New technologies and stricter environmental regulation also tended to reduce the cost advantages of large fossil-fuel steam-power plants.

In the second half of the 20th century, the nation's electric industry

began to feel the strains of growing demand and rising costs. In 1965, New York City suffered a blackout when the utility supplying the city experienced problems at a generating plant and could not obtain power from nearby utilities. To prevent future

blackouts, the nation's utilities formed interconnections to provide back-up sources of power and an organization (the North American Electric Reliability Council) to oversee their coordination. The ability of one utility to tap others through grids (networks of interconnected power plants) and pooling arrangements (joint operating management of multiple independent utility companies) meant that each individual utility no longer needed as large a capacity as it had previously.

Beginning in the late 1960s, advances in operating efficiency that had been achieved almost regularly with the introduction of new large-capacity steam-power plants began to fade. New large fossil-fuel steam-power plants failed to achieve the efficiency in converting heat to

electric power that had been expected of them, and the reliability of these large plants, as well as the reliability of large nuclear steam-power plants, proved to be less than that of small plants. Downtime for maintenance increased for large plants, both fossil-fuel and nuclear, making it difficult for large plants to attain the scale efficiencies they had been expected to achieve through high output rates. Furthermore, some utilities shifted to lower operating rates, running their generators at less than full capacity, to improve reliability and reduce maintenance needs, further undermining efficiency.

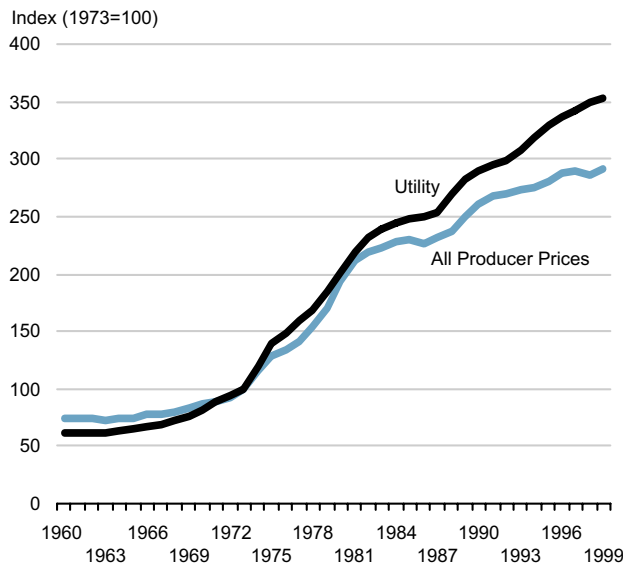
In the 1970s, utilities' costs increased rapidly. Costs of fossil fuels began to rise. Also during this period, construction costs for electric utilities

began to rise, and they rose faster than overall producer prices (Figure 1). Besides actual costs for construction, financing costs rose as construction periods lengthened as a result of the growing complexity of large power plants and longer regulatory reviews during construction. Rising costs led utilities to postpone or cancel plans to build more plants. After the accident at the Three Mile Island nuclear generating plant in March 1979, opposition to nuclear power plants increased and safety regulation expanded, resulting in a drop in nuclear plant construction.

As construction and fuel costs rose in the 1970s, the cost of electricity began to rise sharply after falling during the previous decade (Figure 2). At the same time, growth

FIGURE 1

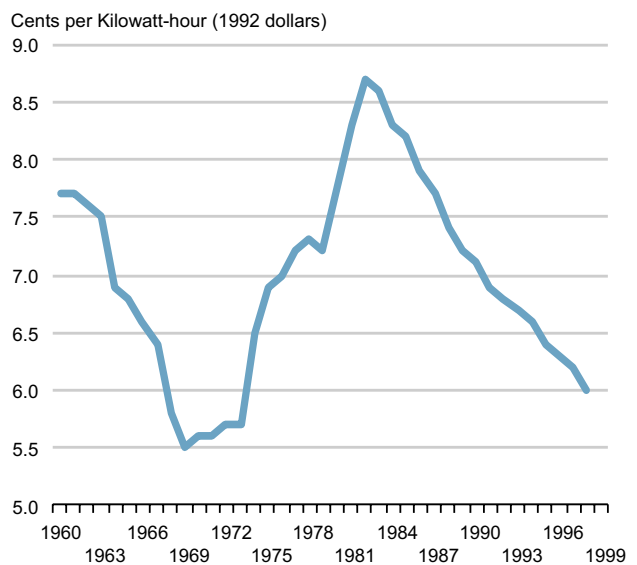
Electric Utility Construction Costs Compared with All Producer Prices*



* Handy-Whitman Index of Public Utility Construction Costs: Total Plant-All Steam Generation (Source: Whitman, Reardon and Associates, LLP)
 Producer Price Index-Finished Goods (Source: Bureau of Labor Statistics)

FIGURE 2

Real Retail Price of Electricity*



* Deflated by the chained gross domestic product price deflator. Source: Energy Information Administration

in the output capacity of the nation's electric power generators (called capability) began to slow (Figure 3). But electricity usage continued to increase despite some dips associated with economic slowdowns (Figure 4). The rising costs of electricity and the concern that generating capacity would not increase in line with growing demand for electric power prompted a search for new ways to meet the nation's electricity needs.

In the 1970s, the design of new generating systems focused on reducing fuel costs, but the new systems also demonstrated that electricity could be produced efficiently on a scale much smaller than that of a typical large steam plant used by electric utilities. In addition to their more economical use of fuel, the

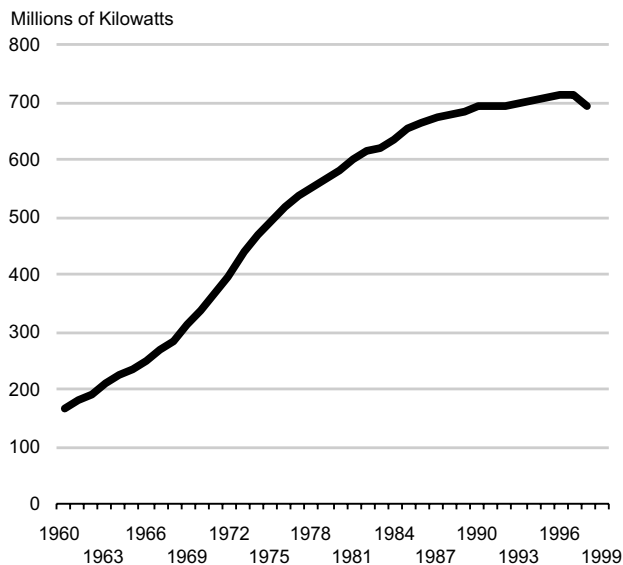
new smaller generators generally produced less pollution and could be built and put into operation more quickly than was possible with the older technology. Two main approaches were implemented. One was more efficient cogeneration technology. In cogeneration systems, heat is produced for an industrial process, and any excess heat is used to drive a turbine to produce electric power, or steam is used to drive an electric generator, and the waste heat from the generator is used in an industrial process. Cogeneration systems are used mainly by large manufacturing firms. The other technological innovation was combined-cycle generating systems, which use waste heat from gas turbines (driving electric generators) to produce steam for steam turbines

(also driving electric generators), thus getting additional electric power from the same amount of fuel. These technological innovations bolstered arguments against the economies of scale model that had motivated the development of large utility firms and their regulation. They also made it possible for large commercial users of electricity to produce their own power or to use the option of producing their own power as a bargaining strategy to obtain electricity at negotiated prices lower than those on existing rate schedules.

As part of the federal response to rising energy costs and slowing expansion of electric capacity, the Public Utility Regulatory Policies Act (PURPA) was enacted in 1978. This law allowed the formation of

FIGURE 3

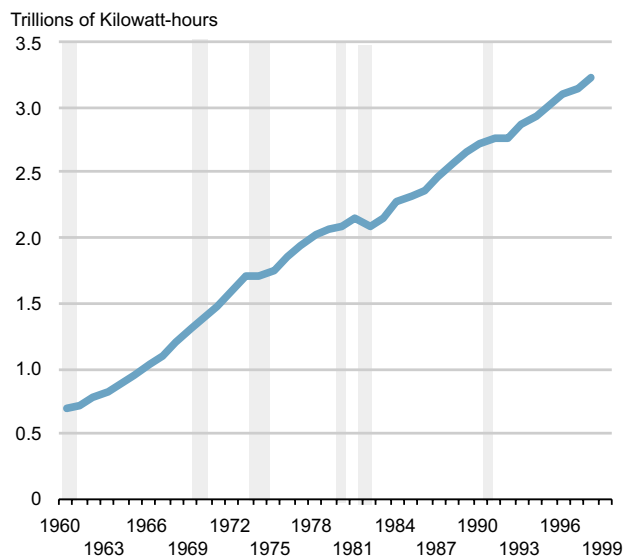
Electricity Utility Capability*



* Maximum continuous output of all generating plants in the U.S.
Source: Energy Information Administration

FIGURE 4

Electricity Usage*



* Total annual usage of electricity in the U.S.
Source: Energy Information Administration
Shaded areas indicate recession years.

companies that could generate electric power without being subject to federal and state regulation as utilities. The law also required utilities to purchase the output of these new suppliers at a price equivalent to the cost the utility would incur if it were to build a new generating plant itself (called avoided cost). Besides the technological factors that influenced the new view of electric markets as legislated in PURPA, new thinking about the presumed efficiency of natural monopolies began to suggest even further steps in opening electric power markets to competitive suppliers.

REFINING MONOPOLY THEORY

Around the same time that questions about the need for and efficiency of large generating plants were being raised, economists began to question the validity of the argument that the existence of natural monopolies inevitably leads to the combination of higher prices and lower output that regulation was intended to overcome. Consequently, economists began to offer explanations of how industries that might be natural monopolies could, in fact, behave competitively.

The main theory of how industries with high fixed costs and economies of scale could be competitive was based on the insight that firms in such industries could compete to be the single supplier of an entire market. Such competition could result in prices and quantities supplied in that market at the same levels as would occur if several firms were competing within that market.¹ Elaboration of this theory posited that even if only one firm established itself as the supplier for the whole market, it could not raise prices to monopoly levels if there remained a threat that a new firm might enter the market. Markets open to entry and exit by potential competitors are said to be “contestable.”² An important element of contestable

¹ An early statement of this theory was by Harold Demsetz, “Why Regulate Utilities?” *Journal of Law and Economics*, 11, April 1968, pp. 55-65.

markets is that competitors’ sunk costs are low even if fixed costs are high. Construction of a large plant is a fixed cost—one that has to be incurred in order for the firm to operate but that does not vary with the scale of production. This fixed cost becomes a sunk cost if it cannot be recovered by the firm, say, by selling the plant to another firm or by switching it to other uses.

Suppose a firm cannot recoup the cost of a plant. Then in determining how much to produce and how much to charge for its product, the firm should ignore these sunk costs. But a firm that is deciding whether to enter the industry must take into account whether it pays to incur the sunk costs of entry. The existing firm that ignores sunk costs can price its product below that of the potential entrant that cannot ignore sunk costs. Under these circumstances, potential new competitors would be reluctant to enter the market. To be contestable, markets must allow firms to enter without incurring large sunk costs in order to begin operations, and they must not allow firms already in the market to have exclusive use of already existing sunk-cost facilities.

The development of contestable markets theory began to influence public policy with respect to regulation in a variety of industries previously considered natural monopolies, notably, air transportation, trucking, and telecommunications.³ Such industries were considered natural monopolies largely because of their high fixed costs. Elements of contestable markets theory, especially the

² A detailed explanation of contestable markets is given in William J. Baumol, John C. Panzer, and Robert D. Willig, *Contestable Markets and the Theory of Industry Structure*, New York: Harcourt Brace Jovanovich, 1982.

³ See, for example, Elizabeth E. Bailey, “Contestability and the Design of Regulatory and Antitrust Policy,” *American Economic Review*, 71, May 1981, pp. 178-83. Besides discussing entry conditions, this article also discusses the need to handle sunk-cost problems by encouraging access to or transfer of sunk-cost facilities.

significance of keeping sunk costs low, provided a rationale for reevaluating the regulation of utilities. For example, the separation of generating facilities from other segments of a utility’s operations, especially transmission facilities, as promoted by the Federal Energy Regulatory Commission’s (FERC) orders and state restructuring laws, is one way of reducing utilities’ sunk costs and making markets more contestable. Prior to these changes in regulations, existing generating facilities were sunk-cost facilities because they could not easily be sold separately (to recover their cost of construction) by the utilities that owned them. With restructuring, generating facilities became more salable, and several utilities have, in fact, divested their generating facilities. Likewise, federal and state mandates on “wheeling” allow competing power suppliers to use the same transmission facilities, which are sunk costs, instead of giving exclusive use of them to a single utility.⁴ Wheeling expands the options available to an electricity distributor: the distributor can obtain electricity from less expensive suppliers to which its own transmission lines are not directly connected.

NEW FEDERAL REGULATORY STRUCTURE

Previously viewed as monolithic natural monopolies, electric utilities have come to be considered integrated firms, combining three stages of the electric power supply system: generation (producing electricity at generating plants), transmission (moving electricity over high-voltage lines from generating plants to distribution nodes), and distribution (moving electricity over low-voltage lines for delivery to the final user). Recent federal regulations have had a major impact on the first

⁴ Wheeling is the transmission of electricity from a first-party producer over the transmission lines of a second-party utility to a third-party utility for final distribution to the consumer.

two stages of the electric power supply system: generation and transmission; the third stage, distribution, has been largely unaffected.

The National Energy Policy Act of 1992 (NEPA) set the stage for major changes in electric utility regulation. NEPA expanded the class of independent firms and subsidiaries of utilities that could be formed to generate electricity without being subject to federal utility regulation. These so-called exempt wholesale generators had first been legalized by the Public Utilities Regulatory Policy Act of 1978, as part of the national government's response to high energy prices in the 1970s. NEPA also directed FERC to require wholesale wheeling. To implement this mandate, FERC issued two regulatory orders in 1996 that brought sweeping changes to the industry.

The first, Order 888, requires owners of all interstate transmission lines to make them available to all power generators under equal terms for wholesale transmission (wheeling). Prior to issuing the order, FERC and the Nuclear Regulatory Commission ordered wheeling arrangements for individual utilities on a case-by-case basis. Most transmission lines are currently owned by utilities, but Order 888 also established standards under which transmission systems may be operated as free-standing entities. Thus, the order laid out a regulatory framework under which the production and transmission of electricity could be conducted by different companies instead of being combined in a single firm, as under the previous structure of regulated monopoly utilities.

The second regulatory order, Order 889, requires every electric utility to provide all other utilities and power providers with online, real-time information about its available transmission capacity. This information provides the basis for spot markets in transmission capacity, making wheeling more flexible in responding to changing needs for electric power in different areas. Order 889 also requires that utilities establish separate

administration and accounting for their transmission and power-generation activities. This provision of Order 889, along with the provisions of Order 888 that set standards for free-standing transmission companies, promotes arm's-length dealings between owners of transmission systems and power providers. By fostering equal access to transmission, these orders reduce the possibility that

enterprise was the most cost-effective way to provide electricity. The new state laws separate the generation stage from the transmission and distribution stages and allow businesses and households to select their own supplier.⁶ At this time, transmission and distribution remain regulated monopolies. Transmission and distribution systems are operated mostly by the former monopoly

The new state laws separate the generation stage from the transmission and distribution stages and allow businesses and households to select their own supplier.

owners of transmission systems that also own generation facilities will discriminate against other power providers. For example, in the absence of a rule such as Order 889, a firm owning both generation and transmission facilities could shift some of its generation costs to its transmission operation, thereby lowering its generation price and raising the price it charges other power suppliers for transmission. Since the states began deregulation, FERC has been studying further changes in regulation to increase competition in the industry.⁵

NEW STATE REGULATIONS

By opening up power production to competition and enforcing open, nondiscriminatory transmission, federal policy established market features that enabled states to open the retail market to competition for selling electric power to consumers. Traditional regulation of electric utilities was based on the notion that the three stages of electric supply—generation, transmission, and distribution—were technically inseparable or that their combination into a single

utilities, although changes are taking place in these sectors as well. (These changes are discussed in the section on the future of the industry.)

The new structure of intra-state electric markets is broadly similar among the states that have instituted competitive electric markets. Firms that meet certain requirements for operating standards to ensure reliability of supply are allowed to offer electricity at unregulated prices. Former monopoly utilities are also allowed to offer unregulated prices for electric power. However, because these former monopolies created systems designed to supply the whole market, they have higher total costs than new entrants. Referred to as "stranded costs," these expenses result largely from the utilities' reliance on large plants or long-term contracts for energy negotiated in the 1970s, when energy prices were higher than they are today. Most state restructuring laws allow former monopoly utilities to recoup these expenses through a charge on all consumers' electric bills, regardless of whether they switch to a new supplier. At the same time, most states require the former utilities to

⁵ Recently, FERC issued Order 2000, calling for further separation of transmission and generation.

⁶ A supplier may be a utility generator, a non-utility generator, or a marketing company that sells power supplied by a generator.

cap or reduce their combined charges for electric generation, transmission, and distribution. Costs for transmission and distribution remain regulated, but they must be separately enumerated on consumers' bills.

DEREGULATION IN THIRD DISTRICT STATES

Changes in federal laws and regulations paved the way for restructuring, but state action is necessary to actually bring about changes in the regulations that govern the electric power industry. As noted earlier, 23 states have passed legislation that now permits or will soon permit retail consumers to choose electric suppliers. All three states in the Third District are among those 23.

Pennsylvania. Pennsylvania was not the first state to implement changes, and it is not the largest state electricity market. But it has become the focus of national attention for its restructuring experience because a greater percentage of consumers there have switched to new electric suppliers than in any other state. Consumer choice was phased in during 1998, and all state residents gained the right to choose their electric supplier in January 1999. The state law authorized the Public Utility Commission to cap each former monopoly utility's total combined charges for generation, transmission, and distribution for four-and-a-half years. (Charges for generation are deregulated, but transmission and distribution charges remain under state regulation.) The seven former monopoly utilities that served the state now compete with 23 electric power suppliers for residential customers and 45 suppliers for commercial and industrial customers.

As of mid-year 2000, the percentage of customers that have switched suppliers from the former monopoly utilities ranged from just under 1 percent for the utility that lost the least customers to 30 percent for the utility that lost the most customers for residential service. For commercial customers (such as stores and offices) that switched, the percentages ranged from 1 percent to 30 percent, and for

industrial customers (such as manufacturers), the percentages ranged from none to 44 percent.

A reason often cited for Pennsylvania's greater participation in choice of electricity suppliers is the state's treatment of stranded costs. When it determined the new pricing structure for generation versus transmission and distribution charges, Pennsylvania's Public Utility Commission set the stranded cost charge that would appear on all consumers' bills at a relatively low amount. This left the former monopoly utilities with the need to recover these costs through

If the current restructuring trend continues, electric power generation and transmission will probably evolve into two distinct industries.

the unregulated prices they would charge for generation. Consequently, new suppliers that were allowed to supply electric power in the state have been able to charge less for generation than the former monopoly utilities. With greater savings possible from switching electricity suppliers, Pennsylvania residents have switched in greater numbers, proportionately, than residents of other states that have enacted consumer-choice legislation.

Another factor that possibly accounts for the extent to which customers have chosen new suppliers in Pennsylvania is the emergence of buyers' consortia. Envisioned by some consumer-advocacy groups as a means for individual customers to combine their buying power, consortia are groups of customers who bargain jointly with suppliers. In Pennsylvania, buyers' consortia do not face the regulatory restrictions that they do in many other states, and Pennsylvania has a well-established tradition of buyers' consortia. In fact, some existing consortia simply added electric

power to the list of products they buy for their members. Consortia simplify the switching process, speed it up, and increase the number of customers who switch. In Pennsylvania, businesses, school districts, municipal governments, and even state government agencies have combined to negotiate contracts with sole suppliers. According to the Pennsylvania Public Utility Commission, consortia members have obtained savings greater than the average available to individual customers under the new law allowing choice of supplier.

A third factor boosting changes to new suppliers in Pennsylvania is the state's extensive consumer-education program. The Public Utility Commission and community organizations have been very active in providing information on consumer choice and instruction in comparison shopping and selecting electric power providers.

Initial price reductions in Pennsylvania ranged from 2 percent to 10 percent, depending on the type of user and the specific provisions of the service arrangements, such as interruptibility. In Pennsylvania, as well as in other restructuring states, the extent of price reductions in the longer term will depend on the balance between the demand for electric power and the number of new suppliers that establish themselves in the market.

New Jersey. New Jersey enacted consumer choice for electric (and gas) suppliers in February 1999, and the program went into effect in November of that year. The law mandated an immediate 5 percent reduction in total charges for generation, transmission, and distribution from former monopoly electric utilities and provided for further reductions up to a total of 15 percent, to be maintained for at least four years. The New Jersey Board of Public Utilities will determine former monopoly utilities' stranded costs and allow their recovery over an eight-year period. As of March 2000, 32 companies had been licensed as energy suppliers in New Jersey, in addition to the four former monopoly utilities operating in the

state. By the middle of 2000, approximately 2 percent of the state's residential electric consumers had switched from their former regulated utility to a competitive supplier and 6 percent of nonresidential consumers had switched.

Delaware. Delaware's electric restructuring law was signed by the governor on March 31, 1999, and will take effect in two stages: large customers could choose suppliers as of October 1, 1999, and all customers can choose starting April 1, 2001. Rates will be reduced 7.5 percent and frozen until September 30, 2002. The new law provides for recovery of stranded costs for the state's sole investor-owned electric utility through a charge applied to large commercial and industrial electricity users; it does not apply to small businesses or residential customers. As of March 2000, the Public Service Commission had certified 16 companies as electricity suppliers, in addition to the former utility.

MORE CHANGES AHEAD?

So far, the restructuring of the electric power industry has been influenced by developments in economic theory, electric power technology, and market structure. Some of these developments have advanced further than others. For example, the recent price spikes in California, a restructuring state, have been attributed in part to the absence of a market in the state that would permit more efficient transactions between power generators and power distributors. Another difficulty facing California, and states in some other regions of the country, is the lack of sufficient capacity in the power grid for wholesale transmission of power into the region at times of peak demand. As the California experience indicates, merely eliminating monopoly among utilities is not likely to provide the hoped-for benefits of a more complete restructuring of the electric power industry.

Despite the elimination of electric power monopolies in many


states, the generation sector of the industry is not yet fully open. State restructuring laws have deregulated only investor-owned utilities and cooperatives. These two classes of utilities together supplied 81 percent of the nation's electric power in 1998. The rest of the national supply comes from federally owned utilities or those owned by state and local governments. Although these two classes together supplied just 19 percent of electric power in 1998, in some regions of the country they are dominant. So far, these classes of suppliers have not been included in state restructuring moves. FERC has recommended that government electric utilities open their transmission systems and be fully integrated into the emerging national market, and several bills to this effect have been introduced in Congress.

The restructuring laws passed by states to date envision the continued operation of transmission and distribution systems as regulated monopolies. However, policymakers have already begun to formulate a structure that at least partially opens up the transmission sector of the electric power industry. FERC Order 888 established standards under which transmission systems may be organized as free-standing entities, referred to as independent system operators (ISOs), clearing the regulatory way for such development. Several ISOs are already in operation.⁷ ISOs are expected to eliminate discriminatory practices by separating management of transmission facilities from their generator-owners and to set prices that avoid an undersupply of power or congestion of the grid.

More recently, FERC has required all utilities that own, operate,

or control interstate transmission facilities to file proposals to create or participate in a regional transmission organization (RTO) that will provide nondiscriminatory access to transmission grids. The RTOs are similar to ISOs, but FERC has set standards for RTOs' independence from power suppliers, as well as for geographic scope, system reliability, and operational authority and responsibility. If the current restructuring trend continues, electric power generation and transmission will probably evolve into two distinct industries.⁸ Many details of grid operation need to be worked out to ensure reliable service and equitable treatment of power suppliers. Transmission organizations, power suppliers, and FERC are giving their attention to these details.

While electricity producers and consumers are adjusting to the recent changes in the electric power industry, a potentially more significant development is looming. Small generators that will be located at the site of use are becoming available for both industrial and residential users. Referred to as distributed power, these small units are dedicated to their owner's or primary user's needs; nevertheless, they can be connected to the electric distribution system. At times when their owner (or primary user) does not require their full output, they can provide power through the distribution system to other users. In this way, the system that used to deliver electricity in only one direction, from the utility to its customers, could become a two-way system.

The past few years have brought great changes in the structure of the nation's electricity system, but the system is still evolving. 

⁷ California utility regulators were the first to approve an ISO, which began operation in March 1998. An ISO covering most of Pennsylvania as well as all of New Jersey and Delaware was started in April 1998. The ISOs that have been established to date are not identical; they differ considerably in organization and operation.

⁸ This structure would be similar to the court-ordered separation of telephone service into distinct local and long-distance markets. See Paul L. Joskow and Roger G. Noll, "The Bell Doctrine: Applications in Telecommunications, Electricity, and Other Network Industries," *Stanford Law Review*, 51, May 1999, pp. 1249-1315.