Business Method Patents and Financial Services

CLIFFORD S. STANFORD

The author is a senior counsel in the legal department of the Atlanta Fed.
This article is an overview of the Atlanta Fed's 2003 Financial Markets Conference, "Business Method Patents and Financial Services," held April 3–5, 2003. The author is grateful to Robert Eisenbeis for shepherding a superb conference on a compelling topic and for inviting him to contribute, and he thanks Bobbie McCrackin for helpful comments on this article.

The modern U.S. economy is experiencing what may be an accelerating shift in the importance of intangible, intellectual, or conceptual assets relative to physical assets. A patent for an Internet-based business method or for a "killer app" software tool can form the basis for an entire enterprise. Today there are highly profitable firms whose assets consist almost exclusively of intellectual property, licensed to generate royalties at virtually no marginal cost of production. Some established firms have discovered that licensing their portfolio of intellectual property assets is far more profitable than producing tangible goods and have modified their entire business strategy as a result.

In keynote remarks at the Atlanta Fed's 2003 Financial Markets Conference, cosponsored with the University of North Carolina School of Law, Federal Reserve Chairman Alan Greenspan outlined dilemmas that "bedevil" economists and jurists alike. Given the increased "conceptualization" of U.S. gross domestic product, and assuming the objective is to maximize economic growth, how does one strike the right balance "between the interests of those who innovate and those who would benefit from innovation"? Does the law correctly calibrate the rewards embodied in intellectual property rights? What are the societal and economic costs of intellectual property rights? Furthermore, does the U.S. system of intellectual property law facilitate a proper delineation of the "metes and bounds" of property rights in ideas?

Greenspan's address provided the foundation for a lively debate among conference participants, who comprised an international mix of economists, legal academics, jurists, policymakers, practicing lawyers, bankers, and technologists. The topic that assembled this diverse group was the emergence and legitimization of "business method" patents in the United States and how this development affects financial services innovation and the future of financial services firms.

Following decades of jurisprudential antipathy to the notion of patenting a method of doing business, the U.S. federal courts raised the flag of surrender in 1998 in the case of *State Street Bank & Trust Co. v. Signature Financial Group, Inc.* The case declared that the mere fact that an innovation is a method of doing business, or is software designed to accomplish business goals, does not mean that, ex ante, such an innovation is not patentable under U.S. law.

This landmark decision, coinciding with the rise of the Internet as a new business channel, provided the impetus for what may be characterized as a new "patent flood." In the years since the *State Street* decision, the volume of patents filed for software and business methods has grown significantly.[1] Given the catalyst of *State Street*, financial services firms realized the potential competitive value of patents on their own business methods and software. Simultaneously, established financial services firms faced the dawning realization that their businesses are not immune from costly patent infringement lawsuits and are threatened by new competitors, including nontraditional players such as technology firms.

U.S. versus International Patent Systems

The conference focused on four key themes.[2] First, now that the last vestiges of subject-matter restraints on business method and software patents have been eliminated, how does the U.S. patent system compare with those of other countries? Professor John M. Conley of the University of North Carolina School of Law noted that U.S. law permitting business method and software patents now appears established and stable. That said, Conley predicted that U.S. courts might begin to erode the enforceability of these patents under the legal standards of "novelty" and "nonobviousness." The courts might thereby seek to stem the tide of overly broad, low-quality, or even spurious patents issued by the U.S. Patent and Trademark Office (USPTO).

Conley then contrasted the patent systems of the European Union and Japan and commented on the impact of GATT's TRIPS Agreement, which took effect in 1995.[3] He noted that the U.S. system relies heavily on dispute resolution before the courts while other systems permit or even encourage the use of administrative procedures before the patent office. The European Union and Japan have technical requirements that appear antithetical to the U.S. approach to business methods and software. Reviewing various case histories, Conley concluded that the differences in these patent systems at the theoretical level are often "not so profound in practice." Further, he found a paucity of empirical evidence to contrast the economic effects of differing approaches to patenting business methods and software.

The Effects on Firms' Business Strategy

How have business method and software patents affected the competitive behavior and strategy of financial services firms? Patents might allow new market entrants and competition, as observed in the biotechnology industry, or they might reinforce the position of established firms, as observed in the semiconductor industry. Professor Josh Lerner of Harvard Business School examined the competitive effects of patenting on the financial services industry, focusing on purely "financial" business methods and on the behavior of investment banks. He estimated that the number of financial patent applications increased three- or fourfold between 1997 and 1999 and that this trend likely has continued. Lerner examined empirical evidence and found that financial patents have been awarded predominantly to large, established U.S. firms. He also found interesting linkages between patenting and firms with ties to the academic community as well as firms focused on debt-related instruments.

Considering anecdotal evidence, Lerner observed that financial institutions now recognize the strategic importance of patent portfolios, both with regard to traditional and nontraditional "paper" competitors. Lerner predicted that while financial services firms have been loath to sue each other and are now building defensive patent portfolios, this stance might break down in the face of a difficult economy or a realization of the licensing value of their patent holdings. Finally, he speculated that in the financial services industry, patents are more likely to help consolidate the position of established firms than to invite new market entrants.

Boon or Bane for Innovation?

Are patents a boon or bane for financial services innovation? Such innovation flourished for decades before *State Street* as a result of incentives other than patent rights. Now that patents have arrived, will the pace quicken, or will patents undermine the system of incentives that drove financial services innovation before the arrival of patents? These and other questions were discussed by Professor Robert P. Merges of the University of California at Berkeley School of Law. Merges identified incentives other than patents that motivate financial services innovation, including "first mover" lead-time advantages, the benefits of "tacit knowledge" not shared with other firms, and attendant reputational advantages as an innovative firm. Even though innovations not protected by patents are subject to reverse engineering and outright copying by competitors, these other incentives have driven significant financial services innovation.

Will patents upset the apple cart? Drawing upon analogies from the nineteenth-century railroad industry as well as today's software industry, Merges found that the financial services industry has responded similarly to the impact of patents by seeking to protect the existing mode of innovation. However, as inferred from the experience of those other industries, the introduction of patents should not damage innovation in financial services. The "codification" of innovation in the form of patents is likely to formalize a previously less formal interchange of innovative ideas and might increase the costs of such sharing in the short term. But Merges posited that this codification will not diminish the beneficial exchange of ideas in the long run and thus will not harm innovation. He also detected salutary effects from the spin-off of innovative firms from established ones and from new innovative firms entering the market using patents as competitive tools. In sum, Merges sees no long-run harm to innovation in the financial services industry resulting from patents but instead some unintended benefits.

The Effects on Policy

Bronwyn H. Hall, professor of economics at the University of California at Berkeley, noted that "most economists view the patent system as a necessary evil: With a patent grant we trade off short-term exclusive (monopoly) rights to the use of the invention in return for two things—(1) an incentive to create the innovation and (2) early publication of information about the invention and its enablement." Given this axiom, what are the implications of business method patents for innovation policy? What policy responses should be considered? Hall asserted that only two things are sure with regard to business method patents—allowing them will result in more business method patent applications, and increased patent activity combined with issuance of low-quality patents will result in an increase in litigation and other transaction costs.

Hall's survey of the literature found that sequential, "cumulative" innovation, which relies upon prior inventions to work, is generally hindered by low hurdles to obtaining and enforcing patents. She concludes that business methods probably fall in this category. A new business method is unlikely to stand alone but is likely to rely upon the prior business innovations of others, which may now be more easily patented. She detected a broad agreement that U.S. business method and software patents have been of low quality as a result of a lack of adequate prior art databases at the USPTO, an overburdened patent office, or permissive "nonobviousness" standards. Low-quality patents increase the transaction costs of innovation, such as litigation costs, and create uncertainty about the risks of innovating in a patent-heavy field.

Hall also surveyed a wide array of policy recommendations to address these issues in the United States. These recommendations range from statutorily reversing *State Street* to raising the bar of nonobviousness standards to providing an improved opportunity for *inter partes* opposition proceedings and reexaminations by the USPTO. The conference participants discussed these policy recommendations but reached no consensus about them.

Conclusions

What lessons can be taken away from the 2003 Financial Markets Conference? Most participants seemed to agree that business method and software patents in the United States are here to stay. Although there are emerging trends detected, and lessons can be drawn from the experiences of other industries, much empirical study remains to be done on patents' effects on financial services innovation, competition, and business strategy. Further, much can be learned through an interdisciplinary approach to the study of these issues. Given Chairman Greenspan's postulation of a shifting economic emphasis to conceptual assets in the modern economy, this conference was indeed "timely and apt."

1. Patents issued by the U.S. Patent and Trademark Office (USPTO) grant the patent holder the right to exclude others from making, using, or selling the patented invention in the United States for twenty years from the date of filing.
2. The conference featured four policy papers as well as four academic papers. Only the policy papers by John Conley and Robert Merges are presented in this issue of the *Economic Review*. For the complete text of all the conference papers, visit the Atlanta Fed's Web site at <www.frbatlanta.org> under "News & Events/Conferences."
3. The General Agreement on Tariffs and Trade (GATT) of 1994 established the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS). The successor to GATT is the World Trade Organization, established in January 1995.

The Uninvited Guest: Patents on Wall Street

ROBERT P. MERGES

The author is the Wilson Sonsini Professor of Law and Technology at the University of California at Berkeley School of Law. This paper was presented at the Atlanta Fed's 2003 Financial Markets Conference, "Business Method Patents and Financial Services," cosponsored with the University of North Carolina School of Law. He thanks Tamar Frankel, Josh Lerner, and the Atlanta Fed staff for help during the preparation of this article.

> Academic research could help to understand whether patenting will encourage or discourage innovation, change the nature of financial innovation, encourage more innovation by smaller players, or change the competitive/cooperative interactions among financial service firms. In part, this yet-to-be-completed work will simply build upon the extensive body of work in the industrial organization field on patenting. However, trying to understand what—if anything—is different about the financial services industry, and the implications for protection of intellection property and the nature of competition, is likely to be a fertile area for future work.
>
> —Peter Tufano (2002, 37)

Up until a few years ago, State Street Bank was just another big bank in Boston. But in 1998 the Federal Circuit Court of Appeals used a patent case filed by the bank to transform the law concerning what is patentable. Since then, the bank's name has been irrevocably linked to a landmark case. Like Linda Brown of *Brown v. Board of Education* fame or Ernesto Miranda, who lent his name to the famous Miranda warning ("You have the right to remain silent . . ."), State Street Bank will be forever associated with a major inflection point in U.S. law.

For many in the financial services industries—banking, investment banking, stock brokerage firms, and the like—*State Street Bank & Trust Co. v. Signature Financial Group, Inc.* was a bolt from the blue. How could patents apply to something as amorphous as the design of a new mutual fund system? Light bulbs, telegraphs, integrated circuits, foolish gadgets like self-tipping hats, maybe, but how could financial products be patentable?[1] As my young son might put it, what's up with that? And more to the point, regardless of where these new patents came from, how would they affect the financial world? Would they help or hurt the financial services industries in the long run? And had anyone thought this all through before making *State Street Bank* a household name outside Wall Street and Boston?

This paper tackles some of these issues. My primary goal is to review what we know about innovation in the financial services industries and to try to discuss intelligently the effect patents will have. But first, as a service to those who might still wonder how these questions got on the agenda, I will try to explain how the patent system got to *State Street Bank* in the first place.

There are two strands to the story: (1) the subversive effects of computer software and (2) the growing fascination with intellectual property generally. I consider each in turn.

The Long and Winding Road to Software Patentability

From the point of view of patent law, the infusion of computer technology has completely changed how the legal system conceptualizes financial services. From a patent lawyer's point of view, many aspects of the financial services industries look like elaborate computer software applications. Despite the differences in climate and dress, Wall Street may as well be Palo Alto, Berkeley, or Redmond, Washington.

After all, one can hear the patent lawyer saying, it's all just software now.

Given this mindset, the patentability of financial services is simply a subset of a larger issue: the patentability of software. This was one of the most troublesome and long-standing issues in patent law for many, many years. Since the early days of the mainframe computer business, when IBM and others tried to get patents on software just as they always had for adding machines and then computer hardware, the patent system tried to grapple with a fundamental conundrum. How could written code—symbols on paper, basically—be a form of technology? Was the patent system of Thomas Jefferson, the MacCormick reaper, Orville Wright, and Thomas Edison the proper home for a series of written instructions to tell a machine what to do?

From a patent lawyer's point of view, many aspects of the financial services industries look like elaborate computer software applications.

The tale of how the patent system stopped worrying and learned to love computer software is a long one. I will hit only the highlights here. After the Supreme Court expressed grave doubts about the whole enterprise in the early 1970s, software went underground in the patent system. It reemerged in the form of patents claiming essentially various pieces of machinery that were assisted by computers running programs (that is, software). Thus, the famous 1980 case of *Diamond v. Diehr* (450 U.S. 175), which upheld the validity of a patent on a rubber-curing machine—a machine that happened to be assisted by a computer running software.

From 1980 until the mid-1990s, patent lawyers pushed the envelope defined by the *Diehr* case. Software was buried in patent claims. Wherever possible, attention was directed to conventional industrial processes that were accomplished using a computer, and the computer just happened to run software. As these inventions were characterized, software was never an end in itself. Yet patent lawyers were forced to resort to ever more creative feats of characterization because software was in fact increasingly separate and distinct from the hardware it ran on. Eventually, the elaborate game of "hide the software in the claims" culminated in a series of claim types. I will explain one of several—the "general purpose computer" claim.

In these claims the invention is described as a general purpose computer, that is, one capable of running many different programs. The claims go on to state that this computer is configured a certain way—configured by software as the computer runs it, that is. Thus, to a patent lawyer, when I shut down my Microsoft Word for Windows application and open Microsoft Excel, I am not just moving in and out of different computer programs. I am creating a new computer! When I open Excel, I am reconfiguring the hardware rather than running a new program.

Although no judge ever actually articulated it, everyone seemed to understand that these characterization games had gotten out of hand. Legal practice did not reflect underlying technological reality. And the computer software industry had simply gotten too big by the 1990s for the patent system to ignore it. Throughout the 1990s there were a series of decisions concerning software that subtly signaled the beginning of the end of many of the old games. Software qua software was no longer strictly forbidden. By the mid-1990s, software in usable commercial forms could be effectively patented.

Despite the sense of change, no single case had clearly stated the end of the old regime. Then along came *State Street Bank*. This case represented a perfect opportunity to clear up any lingering doubts about the patentable status of software. And the Federal Circuit court took advantage, rendering the sweeping opinion now so well known to the financial community.

From the perspective of the history sketched here, then, *State Street Bank* did not come out of the blue—far from it. The decision was the culmination of a very long digestive process. After initially choking on software and then letting only a little bit slip through, in disguise, the patent system finally gave in. Financial services software just happened to be on the menu when the Federal Circuit court got serious about software.

The "Shifting Baseline"—or the Propertization of Just about Everything

I have tried so far in this section to put business methods in the context of the evolution of software patent law. But an even broader change has been taking place, one that is also important for an understanding of how *State Street Bank* came to pass.

Not too long ago, intellectual property scholars could speak confidently of "the competitive baseline"—the idea that property rights were a deviation from commercial norms embodied in our legal system. Patents, copyrights, and trademarks were the exception; open access to rivals' products was the rule. All this has changed in recent years. As I argued in a recent article, the principle of philosopher John Locke—labor yields property—has displaced the competitive baseline:

> The shift that has occurred has taken place at the deepest substratum of the field, down where the foundational principles bump and grind against each other. One massive construct, the principle of the competitive baseline, has started to give way. Under this notion, IP [intellectual property] rights were envisioned as a rare exception. The general rule—the law's deep default—was open and free competition. This was always opposed by a counterprinciple, the idea that labor equals property. On this view, property rights are a matter of desert: in true Lockean fashion, property arises when you mix your effort with the found assets of the natural world. When seen from the perspective of laboring creators, the proper baseline is to protect all manifestations of creativity that take more than a trivial amount of effort. This was a powerful principle, to be sure, but until recently not usually powerful enough. The great tectonic shift of recent years has reversed this, however. Now it often seems as though the labor-equals-property principle dominates. Increasingly, courts and legislators seem to believe that if one type of labor deserves a property right, then others do as well. And so all manner of intangibles meet with protection—even when, in the past, the competitive baseline would have militated against it. (Merges 2000b, 2239–40)

The rise and fall of fashionable ideas is certainly nothing new to the world of finance. One paper on financial innovations is even titled "Boom and Bust Patterns in the Adoption of Financial Innovations" (Persons and Warther 1997). My point here is simply that these are boom times for the concept of intellectual property. Businesspeople, the media, policymakers, and academics all seem fascinated by the idea. It is thus no wonder that, when confronted with a claim to property rights over some novel subject matter, a judge living in this environment is less likely to ask "why?" and more likely to say "why not?" This tendency is a simple fact of our world and no doubt has some influence in cases such as *State Street Bank*.

So where are we now? The table (on page 4) gives us some idea. It presents totals for patents in class 705 of the U.S. Patent Classification system, which is titled "Data Processing: Financial, Business Practice, Management, or Cost/Price Determination," for the years 1994 through 2001.[2]

As with so many things, the numbers tell the tale. Financial innovations are now patentable subject matter. Now that patents are here, the question is, are they really necessary? To answer that, we need to know something about how financial firms protected their investments in innovations before the advent of patents.

The "Appropriability Environment" of Traditional Financial Services Industries

The financial services industries appear to be highly innovative. In the area of traded securities alone, it is estimated that in the 1980–2001 period, the securities industry generated between 1,200 and 1,800 new types of securities (Tufano 2002). Innovation in securities occurs to fill gaps in available instruments. New securities are constantly being devised to shift risks in ways not otherwise possible and to provide payoffs for outcomes that current securities do not cover (what financial economists call "market completeness"). Outside of securities per se, there is no shortage of innovations in the world of finance. New contracts, new transactional technologies such as automated teller machines, and even entire new exchanges have all been common in the past twenty-five years.

1. As many readers will be aware, the *State Street Bank* decision actually goes well beyond financial services. The case authorizes patenting of any "method of doing business" or, more precisely, removes "business methods" from the list of things that are not patentable. In this paper I limit my discussion of *State Street Bank* to its impact in the industry in which it arose—financial services. For more general observations, particularly on the knotty issues of patent quality control the case raises, see Merges (1999).
2. Class 705 is conventionally associated with business method patents even though some relevant patents are found in other classes. The patent at issue in *State Street Bank & Trust Co. v. Signature Financial Group, Inc.,* 149 F.3d 1368, 47 U.S.P.Q.2d 1596 (Fed. Cir. 1998), *cert. denied,* 119 S. Ct. 851 (1999), the case that changed the law in this area, is in this class. See U.S. Patent 5,193,056, "Data Processing System for Hub and Spoke Financial Services Configuration," filed March, 11, 1991, and issued March 9, 1993. Note the issue date—an indication that financial services innovations were finding their way into the patent system even before the practice was explicitly blessed by the Federal Circuit court in 1998.

TABLE

Number of Class 705 Patents Issued

Year	Patents
1994	268
1995	203
1996	274
1997	382
1998	743
1999	1,004
2000	1,062
2001	876

Source: <www.uspto.gov/web/offices/ac/ido/oeip/taf/cbcby.pdf>

Scholars of innovation are well aware that intellectual property rights are not the only mechanism firms employ to recoup product development investments. The general term for this issue in the literature is "appropriability" (Teece 1986). The empirical evidence establishes that patents are considered essential to appropriability in only a few industries—most notably, pharmaceuticals and some branches of the chemical industry (Cohen, Nelson, and Walsh 2000). In other industries, the standard nonpatent appropriability mechanisms include

- lead-time or "first mover" advantages,
- cospecific assets, uniquely adapted for use with the innovation, and
- trade secrecy/tacit knowledge.

In financial services, lead-time, cospecific assets, and trade secrecy/tacit knowledge seem to be important. I consider each in turn.

Cost-saving lead time. In a series of highly illuminating studies, Peter Tufano documented the financial innovation process. Tufano's original paper (1989) studied fifty-eight financial innovations introduced between 1974 and 1986. The innovations were in mortgage-backed securities, asset-backed securities, non-equity-linked debt, equity-linked debt, preferred stock, and equities. These innovations were created almost exclusively by the largest investment banks, with six banks in particular accounting for over 75 percent of "pioneering deals" (Tufano 1989, 219). Large banks were more dominant in innovative deals than in deals overall—making financial innovation very much a game for big players.

Tufano's finding regarding the dominance of large firms in the "innovation game" is echoed by Frame and White (2002):

> For example, casual empiricism leads us to notice that relatively large financial services providers have been important innovators. Merrill Lynch was the developer of the "cash management account"; Salomon Brothers was the leader in developing stripped Treasury securities; the larger commercial banks led in developing and offering "sweep" accounts, ATMs, and Internet transactions for customers. But it would be useful to have a more formal "census" of innovations and their originators and the characteristics of those innovators. (Frame and White 2002, 13, fn. 16)

Tufano studied the appropriability strategies of financial innovators. He found that innovation was indeed costly; he estimates that

> Developing a new financial product requires an investment of $50,000 to $5 million. This investment includes (a) payments for legal, accounting, regulatory, and tax advice; (b) time spent educating issuers, investors, and traders; (c) investments in computer systems for pricing and trading; and (d) capital and personnel commitments to support market-making. In addition, investment banks that innovate typically pay $1 million annually to staff product development groups with two to six bankers. (Tufano 1989, 213)

Tufano finds that investment banks recoup these investments through reduced costs in the market for innovative financial products. The pioneer of a new product has lower costs than its imitative rivals, allowing it to capture a larger market share than imitators. This large market share in turn permits higher profits in the related secondary market for the pioneering product—that is, there are economies of scope. Essentially, even after imitators observe the pioneering product and copy it, the pioneer retains a long-term cost advantage. At the market price set by imitating rivals, the pioneer enjoys "inframarginal costs" and hence supracompetitive profits. Innovators actually charge less than imitators, particularly at first. In addition, a reputation for innovation helps banks in other ways. For example, Tufano describes a class of specialized, client-specific innovations that are rarely imitated (Tufano 1989). In the market to produce these, a reputation for innovation is of course helpful.

This cost-advantage mechanism for appropriating innovation costs is not unknown in other sectors. It seems to explain a good deal of readily copied process innovations in certain industries, for example. The important feature of this appropriability mechanism for our purposes is that it does not rely on property rights to be effective. It does not even rely on informal methods of retaining exclusivity: Everyone in the industry understands that "most new products can be reverse-engineered easily and cheaply" (Tufano 1989, 230). Indeed, rapid diffusion of information about an innovation is actually a marketing advantage for pioneering firms.

Tacit knowledge and reputational advantage. A major area of financial innovation in the past thirty years is securitization, the transmutation of difficult-to-value assets into easily tradable securities. Securitization expert Tamar Frankel has asked why the originators of new securitization practices have not generally sought property rights for them. She begins by noting the difficulty of adapting existing intellectual property categories to the protection of unique securitization ideas. Next, she considers some of the more subtle appropriability mechanisms—tacit knowledge and reputational advantage. Tacit knowledge can be thought of as know-how: the highly detailed, often context-specific knowledge actually required to do a complex job (Polanyi 1967). This knowledge is hard to specify (as more than one artificial intelligence expert can testify), even harder to write down (or "codify"), and harder still to transfer from one person to another (Cowan, David, and Foray 2000). Tacit knowledge is usually therefore defined in contrast to more easily codifiable information.

Frankel argues that tacit knowledge of how to create a novel securitized asset provides a subtle appropriability mechanism to financial innovators:

> Paradoxically, "giving away" an innovation provides many monetary benefits. To begin with, these giveaways may not be complete. Unlike disclosure in applications for patents, disclosures of innovations in advertising, presentations or professional publications are not as complete and detailed. Certain experiences, drawbacks and danger points are likely to be omitted. Some say that following cookbooks of famous chefs rarely seems to produce dishes that taste as the chefs' dishes do. That is not necessarily done by intentionally avoiding an important ingredient from the recipe (although some cooks would be tempted to do so). In a complex area with different actors, it is difficult to transfer fully information in such publications so that the reader can replicate the activity without hands on guidance. Just as the water, cooking utensils, and ingredients may not be identical to those used by the author-chefs, so will the quality of the financial assets, the type of clients and the legal environment of the transactor differ from those of the innovators. These differences may produce difficulties for the novices. (Frankel 1998, 271)

Frankel also provides evidence of reputational advantages accruing to the creators of securities innovations. In this field, lawyers who help transmute illiquid assets into tradable securities make up a small, specialized corner of the legal profession. According to Frankel, "innovators reap the rewards of prestige from enhancing their reputation. For some people, these rewards may be the main driver" (Frankel 1998, 272). This is also consistent with findings by Tufano, who recounts the bankers' view that innovation is the best way to advertise expertise (Tufano 1989, 235).

Intellectual property rights are not the only mechanism firms employ to recoup product development investments. Empirical evidence establishes that patents are considered essential to appropriability in only a few industries.

While one case does not make a trend, a recent trade secret case indicates that appropriability mechanisms other than lead time may occasionally be important. In 1995 Morgan Stanley submitted a proposal to the state of California in response to an unusual request. The state was looking for innovative approaches to securitizing the risks associated with earthquake losses, an insurance market that the state had recently entered in response to perceived market failure in the private insurance business. Investors Guaranty Fund, Inc. (IGF), is a small firm that specializes in coming up with securitization concepts and helping large investment banks to implement them. IGF claimed that Morgan Stanley's submission to the state was based on IGF's "total integrated system" for securitization of insurance risks. IGF had, it argued, successfully employed this system in other securitization projects in conjunction with other banks.

The trade secret suit was dismissed.[3] The court stated that the IGF system was based on public

3. *Investors Guaranty Fund, Ltd. v. Morgan Stanley & Co., Inc.,* 50 U.S.P.Q.2d 1523 (S.D.N.Y. 1998).

domain concepts and was not in fact proprietary to IGF. The court also ruled that the system did not confer a competitive advantage on Morgan Stanley because the state terminated the securitization experiment and implemented a more conventional reinsurance scheme instead.

Industry appropriability and the prior user defense to patent infringement. Good evidence exists that the financial services industry sought to protect established appropriability practices in the wake of *State Street Bank*. Financial services firms lobbied for and obtained a limited defense to infringement that is now part of the U.S. patent statute. Under this "prior user right," firms that have developed and implemented secret internal methods of doing business may not be precluded from using them by later inventors who obtain a patent. A special provision was required to secure this result, as generally U.S. law disfavors a secret prior user compared to a later user who files a patent application.

Long after the advent of the property-rights revolution in science, pure academic research—and the open, property rights–free exchange of information it depends on—continues to thrive.

Prior user rights are common in other countries, particularly in Europe. They provide a measure of protection for firms that develop innovations but do not wish to patent them. They insulate earlier developers from the very expansive reach of property rights granted to later inventors. Many commentators, drawing on the empirical evidence concerning the centrality of trade secret protection as an appropriability mechanism in some industries (Cohen, Nelson, and Walsh 2000), have argued in favor of a general prior user right under U.S. law. But the actual law enacted in the wake of *State Street Bank* is much more limited: It protects only prior inventors of "a method of doing or conducting business" from infringement liability.[4]

Lawyer/lobbyists for the financial services industry very likely drafted this provision—a common occurrence in intellectual property legislation, as elsewhere.[5] In addition, industry representatives also appear to have drafted comments to be entered into the *Congressional Record* under the names of lawmakers from New York and New Jersey—Wall Street territory. These comments provide helpful insight into the perceived threat posed by the *State Street Bank* decision. Thus, the Senate version of the *Congressional Record* includes this entry from Senator Charles Schumer:

> The first inventor defense will provide the financial services industry with important, needed protections in the face of the uncertainty presented by the Federal Circuit's decision in the *State Street* case. . . [T]his decision has raised questions about what types of business methods may now be eligible for patent protection. In the financial services sector, this has prompted serious legal and practical concerns. It has created doubt regarding whether or not particular business methods used by this industry—including processes, practices, and systems—might now suddenly become subject to new claims under the patent law. In terms of everyday business practice, these types of activities were considered to be protected as trade secrets and were not viewed as patentable material (*Congressional Record* 1999b).

The identical statement was entered under the name of Representative Jerrold Nadler (*Congressional Record* 1999c). And a similar comment was entered by Senator Robert Torricelli, who states that "without this defense, financial services companies face unfair patent-infringement suits over the use of techniques and ideas (methods) they developed and have used for years" (*Congressional Record* 1999d).

As Senator Schumer is quoted as saying, financial product innovations have traditionally been "protected as trade secrets." Based on what we know, lead time and reputation might be added to the list. The point of the legislation is to defend these traditional mechanisms against the onslaught of patents. Because of certain technical features of the defense, however, it is not clear that the defense alone will protect financial services firms from the patents of "outsiders." This uncertainty explains why large Wall Street firms are at the same time beginning to acquire some patents of their own.[6]

Property rights enforcement and information sharing in "traditional" areas of innovation. One crucial point of importance at this stage of the discussion is to note that not all property rights are enforced. This concept is often lost on critics of property rights, who positively thrive on presenting and embellishing a gruesome "parade of horribles." With proliferating property rights, we are told, businesspeople could no longer do many things they are accustomed to doing. Every patent owner could prevent everyone else from using their patented technology. And because they could, we are told, they would. Does this claim hold up based on what we know about other fields where intellectual property has arrived suddenly on the scene?

In a word, no. One example comes from academic science. Here open exchange of research findings was long thought to serve as a model of information dissemination in the absence of property rights. Many observers thought the sudden advent of patents on the fruits of basic scientific research—particularly in the life sciences—was sure to kill the scientific enterprise or at least inflict a mortal wound. But it did not. The reason was that although scientists (and particularly the research universities that employ them) aggressively acquire property rights, they almost never assert them against other scientists engaged in academic research. A scientist who draws on the work of peers in doing his or her own research follows a well-understood norm in the field: Patents are asserted only against commercial entities. Fellow scientists operating within the same research community are off limits. In effect, there is an inner circle within which property rights are mutually waived. They are only deployed against private firms operating in the outside circle of the corporate biotechnology industry. Even though many academic scientists work across both circles on a regular basis, they recognize that property rights are appropriate only in the outer circle. Patents are checked at the door when a researcher enters the domain of pure research. These circumstances are why, long after the advent of the property-rights revolution in science, pure academic research—and the open, property rights–free exchange of information it depends on—continues to thrive.

A variation on this theme involves cooperative cross-licensing. In some industries, most notably semiconductors, firms aggressively acquire patents. But they are not typically asserted against commercial rivals in litigation. Instead, firms cross-license large patent portfolios. Sometimes two evenly matched firms cross-license with no royalty payments. For technologically unequal trading pairs, lump sum payments or ongoing royalties change hands. In either event, patents serve as bargaining chips in an elaborate industry scheme of information transfer. Patents mediate, rather than obstruct, the flow of information.

Would patents lead to continued exchange in the financial services industries? It is hard to say. There is some indication that little has changed in the wake of the *State Street Bank* decision. Perhaps the large firms continue to share information amongst themselves, banking patents only as a hedge against outsiders' attempts to use patents to hold up existing firms. And lobbying for a "prior user right" exception to infringement (see the earlier discussion) hints that financial firms' main goal in the post-patent era is to make the world safe for their existing practices. So perhaps the free exchange of information about new innovations will continue for the most part.

Past Responses to the "Patent Plague"

Wall Street's reaction to the threat of patents runs contrary to the simplistic theory of incentives inherent in the patent system. But there are other cases in which an industry has greeted the introduction of patents as more of a threat than an incentive. It may be instructive to review several of these episodes, with the goal of determining how serious the patent threat turned out to be and how effective industry responses were.

Nineteenth-century railroads. The first brief study may seem to come from far afield—temporally and conceptually. But in many ways, the coming of patents to the railroad industry in the nineteenth century looks very like the post–*State Street Bank* world on Wall Street. So far, financial firms have undergone the same shock and surprise that the railroads experienced when they first came to grips with the disruptive effects of patents on established routines of innovation. And Wall Street has responded the same way, though much more quickly—with an aggressive counterthrust to the legal system's incursion into familiar turf. As with the railroads, financial firms have lobbied for legislation to overturn the most damaging aspects of the new patent regime. Indeed, judging by results, Wall Street's response has been more effective so far; the railroads never did succeed in getting favorable legislation passed. By

4. 35 U.S.C. § 273(a)(3) (2002). For more detail, see Merges and Duffy (2003, 172–73).
5. For a limited defense, see Merges (2000a) (reviewing literature on alternatives to rent-seeking and capture theories of lobbying). It should also be noted that the sponsor of the bill that included what is now section 273 of the Patent Act stated that this provision was not intended solely for the benefit of the financial services industry: "The earlier-inventor defense is important to many small and large businesses, including financial services, software companies, and manufacturing firms—any business that relies on innovative business processes and methods" (*Congressional Record* 1999a).
6. For example, in December 2002, CitiCorp had twenty-eight patents, and Merrill Lynch had twenty-six.

contrast, the railroads slogged things out in the legal trenches for many years before beating back the most threatening aspects of the legal onslaught. Despite the differences, there is much to gain in a quick overview of the patent episode in railroad history.

To begin, there was a great deal of similarity in the way innovation progressed in nineteenth-century railroading and in late twentieth-century Wall Street. Innovation in both industries was an inside job: It was dominated by large, vertically integrated firms (Usselman 2002). Nineteenth-century railroads not only laid track and scheduled shipments but also performed service on and made routine improvements to locomotives, switching technology, rails, and all other aspects of railroad technology. Moreover, innovations diffused rapidly to rivals, and this occurrence was an accepted part of the business. Far from preventing this flow of information, the chief technology players at the major railroads saw themselves as part of a larger, cross-firm enterprise. They shared a common culture that included an implicit norm regarding new techniques: I share with you, you share with me (Usselman 2002). There was pride in an innovation that others could use, perhaps even some increment to firm or individual reputation.

There was a great deal of similarity in the way innovation progressed in nineteenth-century railroading and in late twentieth-century Wall Street. Innovation in both industries was an inside job dominated by large, vertically integrated firms.

The "appropriability regime" was dominated by complexity and capital constraints. Locomotive technology, for example, was simply too complex for many firms to get into the industry. There were few rivals around that could gain much from learning about an innovation. New technology alone was rarely seen as conveying a competitive advantage. Reaping the rewards from it required access to the wide array of cospecific assets making up a full-service rail line. Property rights played a very small role in such a setting.

All this began to change by the 1870s. This era saw a host of outside inventors descending on the railroads. They promoted a long series of improvements and enhancements, some centering on safety devices invented in response to highly publicized rail disasters. But many came from mechanics and tinkerers of all varieties, swept up in the fascination with rail and steam that (then and now) seems to hold many in its thrall. The number of patents awarded for various aspects of railway technology grew steadily throughout the nineteenth century (Schmookler 1967).

A modest number of outside inventions were adopted by the railroads during this period. But the patent system really burst into prominence when courts began awarding huge damage awards to the holders of patents who had sued the railroads.[7] In the wake of several much-discussed infringement suits, patent matters rose to the highest levels of discussion within the railroad companies. Although the corporate response took some time to coalesce, by the 1880s the industry was fully mobilized. Two large industry organizations supervised and carefully monitored the progress of important infringement suits, including several at the Supreme Court.

Meanwhile, a legislative response took shape. Railroad executives lobbied hard in congressional hearings against the extension of patents that had been costly to the industry. Lobbying also centered on a bill to overturn a particularly costly doctrine that had arisen in the courts. The "doctrine of savings" used a firm's estimated cost savings due to the use of a patented device as the basis of damage calculations. In the hands of a sympathetic judge or jury, it could lead to very expensive judgments. The industry labored to pass a bill to overturn the doctrine—and very nearly succeeded. But when the Supreme Court in 1878 adopted a more favorable interpretation of the savings doctrine, the industry finally backed off.[8]

Apart from an increase in lobbying expenditures, did the introduction of patents affect the railroad industry? In particular, did the introduction of patents in any way slow down the course of railroad industry development?

The answer is clearly no. Jacob Schmookler documented railroad industry investment, additions to railroad track mileage, and stock prices for the period 1837 until 1950. All three measures showed robust increases throughout the nineteenth century (Schmookler 1967, 116). Of special note is the fact that particularly sharp increases in these measures were recorded at the same time patents were arriving as a major force on the railroad scene (roughly, between 1860 and 1890). Whatever the effects of patents on the railroad industry, they did not bring it to a halt. Of course, growth might have been even more robust in the absence of patents. But, realistically, they did not appear to slow the development of this industry in any significant way.

U.S. software industry. The U.S. software industry voiced very similar concerns when software patents became a reality in the 1980s. Cries were heard throughout the community of computer programmers that patents would kill the goose that had laid the golden egg of software creativity in the United States (Merges and Duffy 2003, 196–203). A particular concern was that software patents would give an advantage to large firms, in particular IBM; there was fear over the clash of a "patent culture"—with its attendant high overhead costs—and the freewheeling and productive culture of programmers who were said to write code not strictly for profit but for technical sophistication and elegance.

A funny thing happened on the way to the demise of the software industry. It never happened. Standard-setting organizations ameliorated some of the problematic effects of having multiple components of complex software products and protocols owned by separate firms. Several early test cases found the courts being quite reasonable about scope and validity issues with respect to computer software. And most telling of all, programmers forming start-ups found that venture capitalists placed a premium on companies with a robust patent portfolio. So leading-edge firms such as Inktomi moved quickly to establish effective patent portfolios. One reading of the history here is that software entrepreneurs found that patents were decidedly not just for the big guys. In any event, the industry continues to move ahead despite—and in some cases even perhaps because of—the advent of patent protection.

On the other hand, software patents have not changed many of the basic features of the industry, including the importance of "network effects" to many of its products (Saloner and Shepard 1995). Perhaps there is a deeper path dependency in industrial development than we are aware of. An industry, once started on a patent-free basis, establishes an innovation path that later proves relatively impervious to the imposition of patents. Perhaps patents overall simply do not affect the big variables of economic life—industry structure, the basic pace of innovation, etc.—in such an industry to any great extent. While these are somewhat humbling thoughts for a scholar who places the patent system at the center of the economic universe, the historical case studies certainly support such a view. Apart from their role in fostering outside entry, and perhaps a marginal but significant role in making old industries safe for small, entrepreneurial firms, patents do not seem to have shifted the basic parameters of innovation in either railroading or software. If this pattern holds true, we may predict that patents will not significantly affect the overall structure or innovativeness of the financial services industry. To sound a Chandlerian theme: While patents may play a key role in individual firms' strategies, they may not have much impact on industry structure.

Property Rights and the Market for Financial Technology

Research on the emergence of markets for technology may have something to teach here as well. According to this literature, active interfirm markets for technology are increasingly popular for a number of reasons. The major factors are (1) increasing creativity in "mining" intellectual assets for profit, (2) reduced fear of selling ideas to major competitors, and (3) improving and expanding know-how about how to propertize and value intellectual assets (Arora, Fosfuri, and Gambardella 2001; Davis and Harrison 2001).

Viewed from the perspective of this literature, one interesting question is what effect patents will have on formalizing the exchange of information about financial services innovations. In the past, this information diffused out from innovators to other firms in the relatively closed circle of experts in each area.[9] Now, with the advent of patents, these innovations can be (to use the language of economists who study information transfer) codified. Patents play a role here in helping identify discrete units of information for transfer. They also facilitate valuation by clearly demarcating the boundaries of a discrete idea and by feeding into a system of legal and technical experts who specialize in valuation.[10]

7. See, for example, *Chicago & N.W. Railway Co. v. Sayles*, 97 U.S. 554 555–556 (1878) (summarizing district court proceedings from 1865 through 1875); *In re Caewood Patent*, 94 U.S. 695 (1876) (concerning patent for "swedge block" used to repair and straighten worn railway rails).
8. *Chicago & N.W. Railway Co. v. Sayles*, 97 U.S. 554 (1878) (reversing lower court opinions and reining in "doctrine of savings").
9. One piece of evidence from a theft of trade secret case involving techniques for securitization suggests that some explicit information transfers have taken place under the rubric of trade secret licensing. See *Investors Guaranty Fund, Ltd. v. Morgan Stanley & Co., Inc.,* 50 U.S.P.Q.2d 1523 (S.D.N.Y. 1998): "Plaintiff contends that five . . . banks—First Boston, Goldman Sachs, Donaldson Lufkin & Jenrette, Salomon Brothers, and JP Morgan—had received information from IGF about its system under 'confidentiality, proprietary, trade secrets acceptance conditions.'" The case was dismissed anyway on the ground that the plaintiff had not adequately backed up its assertions in this respect.
10. Embodying technical information in a formal property right such as patent can significantly lower the cost of exchanging it with another firm (Arora and Merges 2001).

Patents can therefore push information exchange from an informal basis to a more formal one. Whether this is beneficial depends on the number of transactions that result under each of the two regimes. Currently, information about financial services innovations diffuses rapidly—through informal contacts among the principal designers of innovations, trade press articles, simple observation of what competitors are doing, etc. These information exchanges are easy to miss as they involve essentially zero transaction costs. Every time a businessperson learns something about a competitor's new practice in some area, after all, information has been transmitted.

Apart from their role in fostering outside entry and perhaps in making old industries safe for small, entrepreneurial firms, patents do not seem to have shifted the parameters of innovation in either railroading or software.

What happens when information such as this is propertized—when an intellectual property right (IPR) attaches to it? Total transactional volume may well be affected. But how?

If a sizable proportion of the information is suddenly covered by a property right, the flow of information may well decrease at first. What had been essentially free is suddenly more costly; information acquirers move up their demand curves. Over time, however, a number of offsetting gains might compensate for or justify this additional cost. A bedrock assumption of the intellectual property system is that certain information will not be produced without the special incentive of a property right. Thus, the addition of property rights to the equation will—in theory at least—call forth new and greater creative efforts, resulting in a larger number of innovations. True, some transactions that would have been free will now cost more. But the conventional wisdom from inside the IP system would predict a net increase in innovations. To put it bluntly, there is a possibility that while free transfer of ideas to competitors will end, a robust market in the formal exchange of new financial innovation ideas will lead to more exchanges of more valuable information.

Spin-offs. A related possibility involves spin-offs. Because much of the know-how associated with financial innovations currently resides in large firms, the people to staff new entrant firms will likely come largely from the established players. We are all familiar with many cases of start-up companies emerging from the ranks of established players. The dynamic nexus of restless entrepreneurs, venture capitalists, and corporate lawyers is an important component of the institutional infrastructure of Silicon Valley and other innovation-rich regions. Established firms, confronted with this reality, have responded in recent years by saying in effect, "If you can't beat them, join them." The result is a greater number of spin-offs.

Spin-offs could become an important part of the scene in financial services for a number of reasons. In financial services, broad expertise is required to innovate, at least in some areas. So innovation begins in many cases in large firms. In the language of appropriability, access to the cospecific assets of a large, integrated firm is essential for successful innovation.

But once an innovation is made, there may be reasons why a separate firm makes a better home for it. First is the simple fact that huge, integrated firms may not reward the development of the innovation as directly or effectively as small, highly focused firms do. This "incentive intensity" effect is a well-known advantage of small start-ups. It explains why start-ups often push more aggressively to expand applications of their basic technology into markets far afield from the business of the parent (see the eSpeed story on page 11). Second, in some cases rival firms are far more likely to do business with a small separate entity than with a division of a large integrated rival. When a sophisticated technology-intensive input is being supplied, the buyer may have to reveal sensitive information about its product design or operations. A company may be reluctant to share this information with a direct competitor. This logic seems to be at work at times in the chemical industry, where sophisticated process technologies owing their origins to large, integrated chemical firms are sometimes spun off into independent start-ups (Arora and Merges 2001).

Patents appear to play an important role in spin-offs in some industries such as specialty chemicals (Arora and Merges 2001). Without patents, the risk that the technology will be copied by the spin-off firm's customers is too high. While trade secrecy is a common appropriability mechanism for established chemical firms, spin-offs by definition lack the cospecific assets necessary for a trade secret–oriented strategy to be effective. The only answer is to have strong patent protection.

Is this model possible in financial services? Much depends on the extent to which independent firms can find a market for new financial product and service ideas. If the transaction costs are too high for deals involving these "goods," independent firms will not be viable—regardless of presence or absence of property rights. Markets for pure, disembodied ideas are, after all, fairly rare. Another consideration is whether independent firms can devise and develop enough of these ideas to remain viable. Perhaps it requires access to many operational details and many different professionals to devise new financial products and services. The dearth of financial idea start-ups to date certainly suggests as much. If financial idea start-ups face the problem of a dry product development pipeline, they will not be viable.

Perhaps the Cantor Fitzgerald spin-off eSpeed is an indication of things to come.[11] eSpeed develops and sells pricing and trading software for various securities markets. It started in the bond market, of course, where Cantor Fitzgerald was and is a major player (despite the efforts of terrorists). Building on Cantor's original $200 million investment in new trading technology, eSpeed is branching out into other markets: energy, bandwidth, futures, telephone minutes, etc. (see www.Cantor.com). It appears that eSpeed is serious about research and development, according to a recent 10-K filing:

> We devote substantial efforts to the development and improvement of our electronic marketplaces. We will work with our clients to identify their specific needs and make modifications to our software, network distribution systems and technologies which are responsive to those needs. We are pursuing a four-pronged approach to our research and development efforts: (1) internal development; (2) strategic partnering; (3) acquisitions; and (4) licensing. We have approximately 150 persons involved in our internal research and development efforts. . . . We are continuing to develop new marketplaces and products using our internally developed application software having open architecture and standards. In addition, we have forged strategic alliances with organizations such as Sungard/ASC and QV Trading through which we will work to develop sophisticated, front-end trading applications and products. We expect to license products from and to companies. . . . (ESpeed 1999 Form 10-K, available at <www.sec.gov/Archives/edgar/data/1094831/0000889812-00-001393-index.html> at 42).

At the same time, eSpeed is also a fairly intellectual property–intensive firm, according to a 10-K filing:

> We expect to rely primarily on patent, copyright, trade secret and trademark laws to protect our proprietary technology and business methods. Our license with Cantor includes four issued United States patents as well as rights under domestic and foreign patent applications, including foreign applications currently filed by Cantor (ESpeed 1999 Form 10-K, available at <www.sec.gov/Archives/edgar/data/1094831/0000889812-00-001393-index.html> at 8–9).

And, to the extent the trade press can be believed, the firm has aggressively pursued markets far distant from Cantor's home base of bond trading (*Red Herring* 2000). Indeed, its efforts to enforce some of its patents have brought some criticism already.

Start-ups, or "Silicon Valley comes to Wall Street." Peter Tufano asks whether financial services patents will "encourage more innovation by smaller players" (2002, 37). This section explores the possibility that the answer might be yes—that apart from spin-offs, true start-ups may become a more common sight in financial services.

To a large extent, a long-time observer of the patent system cannot help notice that the best justification—and sometimes, to be truthful, the only one—for the system appears to be to promote the financing of dynamic new entrants. The connection between patents and venture capital financing is a well-accepted part of Silicon Valley practice, though economists are just now taking at a stab at explaining why (Gans and Stern 2002; Hellmann and Puri 2000).

Scholars operating in the tradition of Joseph Schumpeter have made connections between entry by start-up firms, patent protection, and industry structure and competition. Just as Merges and Nelson (1990) argue that multiple, rivalrous sources of innovation often promote faster economic growth, Boot and Thakor (1997) model how different institutional structures might lead to different levels of innovation. They predict less innovation in a financial system of universal banking, especially where it involves significant market concentration. On the other hand, where commercial and investment banking are functionally separated, Boot and Thakor predict more innovation. As with Merges and Nelson, the basic idea is that competition yields increased innovation.

11. eSpeed commenced operations on March 10, 1999, as a division of Cantor Fitzgerald Securities. In December 1999, eSpeed was spun off from Cantor Fitzgerald in an initial public offering (see <espeed.com/about_espeed/history.html>).

It is too early for a systematic test of these concepts. But some intriguing possibilities for the future are suggested by firms exploring the start-up/patent orientation in financial services.[12]

One such firm is Financial Engines, Inc., a Silicon Valley start-up, with its headquarters in Palo Alto and backing from a number of prominent venture capital funds (see www.financialengines.com). Financial Engines makes a business of providing sophisticated, automated on-line investment advice for various investors, typically employees of large companies that subscribe to its services. It services dozens of clients that employ thousands of employees. Notable for our purposes is the fact that Financial Engines has a patent-intensive strategy. As of fall 2002 the firm held five U.S. patents.[13] It also partners with other firms by licensing its financial advice software systems as components in larger investment services packages.[14]

Research suggests that patents may influence not only the overall rate of innovation but also the sources of innovation and, through this, perhaps even industry structure.

Another firm with a similar profile is FolioFN, which permits institutional and individual investors to put together customized investment portfolios including fractional shares of various investment instruments. This approach brings the benefits of diversification to a broader market and deepens the degree of diversification possible with a given investment amount. The FolioFN approach is based on a series of patents, including U.S. Patent 6,338,047, "Method and System for Investing in a Group of Investments that Are Based on the Aggregated, Individual Preference of Plural Investors," issued to Wallman, et al., January 8, 2002. As with Financial Engines, the FolioFN business model requires partnering with other firms to broaden the business, particularly individual and institutional investment advisers.

Patents, contracts, and the viability of start-ups. Both start-ups described in this section plan to rely on partnering. Recent research teaches that patents may play a role in facilitating technology- or information-intensive transactions such as these (Arora and Merges 2001; Hall and Ham-Ziedonis 2001). If this research is accurate, it suggests that patents may influence not only the overall rate of innovation but also the sources of innovation and, through this, perhaps even industry structure. The basic idea in this literature is that property rights can make small entrants viable at the margin in settings where entrants without property rights rarely survive. Hall and Ham-Ziedonis (2001), for example, study the emergence of small "design boutiques" in the U.S. semiconductor industry. This industry is characterized by very large, vertically integrated manufacturing firms. The small entrants gain access to necessary manufacturing assets by licensing their designs—which is possible only in the presence of strong patents, given the strong probability that manufacturing firms could easily copy expensive designs. In the language of appropriability, patents facilitate contractual access to cospecific assets. The general phenomenon is modeled by Arora and Merges, who also describe a case study drawn from the biotechnology industry. There, a supplier of sophisticated inputs used in the manufacturing of biotechnology products survives and thrives dealing with customers whose expertise and know-how would make it easy to copy its "crown jewel" technology. Again, broad patent protection is the key.

It is impossible to say at this point whether financial services patents will permit the emergence of similar success stories. But the fact that experimentation along these lines may already be beginning is intriguing. Together with the eSpeed case study, these start-ups show that patents in the financial services industry have the potential to increase the diversity of organizational forms available to innovating firms in this industry.

Conclusion: Patents and the Ecology of Wall Street

To calibrate the impact of patents on financial services with any degree of precision is not possible. There will be upheavals—patent lawsuits that roil the industry, announced patent grants that trouble industry leaders and threaten established firms and practices, and an overall concern that patents have changed old practices in unwelcome ways.

But beyond this, in the long haul, I will venture a prediction: Patents will not cause any real and lasting problems. I offer this assessment based not on hard empirical predictions but on two detailed historical case studies, one from the nineteenth century (the railroad industry) and one from recent times (the software industry). I chose them because in both industries the adjustments to patents followed the same general pattern. And in both, early concerns that patents would fundamentally undermine innovation were proved quite wrong.

Wall Street did not need patents. It certainly did not ask for them. Innovation was flourishing without them. And when they came, these strange "incentives" were greeted with skepticism, akin to the Reagan-era joke, "We're from the government. We're here to help."

But now they are here. What will happen? The early fear was that they would upset the natural ecosystem that had evolved without them. Like a civilization cut off from the outside world, Wall Street would suddenly be infected with a novel pathogen. There would be sickness where there had been health and balance.

A patent-related epidemic may appear in Wall Street's future. But I doubt it. The industry-backed prior user rights exemption was an early inoculation. And the industry immune system is less likely to be surprised now: Firms are more aware that they need to be vigilant in watching what issues from the Patent Office and in acquiring some defensive patents of their own. Some high-profile patent infringement lawsuits will probably be filed, but a wholesale blindside of the industry appears less and less likely.

At the same time, some unintended benefits may flow in the wake of patents. Perhaps a few new entrants will be viable that would not have been. Perhaps patents will call forth some extra efforts at innovating in some sectors. Stranger things have happened.

Even if not much good comes of it, Wall Street ought to pause before criticizing the advent of patents. Perhaps in an ideal world, policymakers would have studied the financial services industry carefully for a decade before extending patent protection to financial innovations. Hearings would have been held, fact-finding missions conducted. No surprises would have been sprung on an unsuspecting industry by an outsider court with no Wall Street bona fides. The whole exercise would have been much more rational, premeditated, and predictable.

But, as the *State Street Bank* decision demonstrates, that's not how it works in our system. Because our judges are totally independent, they did not have to worry about upsetting Wall Street. And the separation-of-powers principle enshrined in our Constitution means that the Federal Circuit court did not need Congress's permission or the president's blessing to throw a monkey wrench into the operations of a major U.S. industry. The court followed the logic of its own area of expertise and in so doing upset received practices and conventional wisdom. Meanwhile, Congress did not have to clear it with the court when it passed the prior user rights exemption. This sort of institutional dialectic of challenge and response, this series of random outside shocks, is often unsettling at first. Yet it gives our economic and political system vitality, energy, and even (am I really writing this in an academic paper on financial services patents?) a sense of adventure. Ecologists and students of evolution often talk of the beneficial effects of random shocks in the natural world. Perhaps Wall Street ought to pause before criticizing this one. Something good may come of it. In the meantime, old practices will have to be examined. Implicit routines will have to be made more explicit, received wisdom questioned. This shakeup may not be all bad. After all, nature teaches that regular events like this are good—that the uninvited guest is sometimes the most interesting one of all.

12. By some accounts, start-up activity in this area appears to be on the increase. See Heaton (2000), which states, in discussion of a particular start-up, that "many other financial patents are held by similarly situated start-ups and entrepreneurs."
13. See, for example, U.S. Patent 6,125,355, "Pricing Module for Financial Advisory System," issued to Bekaert et al. (patent providing a single pricing module that models both fixed-income securities and equity securities into the future in an arbitrage-free model), and U.S. Patent No. 6,292,787, issued to Scott et al., September 18, 2001, "Enhancing Utility and Diversifying Model Risk in a Portfolio Optimization Framework."
14. See, for example, Tom Lauricella, "State Street, Citigroup Venture to Give Advice on 401(k) Plans," *Wall Street Journal*, June 10, 2002: "For the first time, investors in some 401(k) retirement plans soon will be able to get advice to buy or sell specific investments through the financial-services company administering their accounts. Citistreet, a joint venture of Citigroup Inc. and State Street Corp. that is one of the largest retirement-plan providers, announced the service Monday. Advice provided to investors in the Citistreet plans will be based on analysis and recommendations from Financial Engines Inc., an independent investment-advisory firm."

REFERENCES

Arora, Ashish, Andrea Fosfuri, and Alfonso Gambardella. 2001. *Markets for technology: The economics of innovation and corporate strategy*. Boston: MIT Press.

Arora, Ashish, and Robert P. Merges. 2001. Property rights, firm boundaries, and R&D inputs. SSRN Electronic Library. <ssrn.com/abstract=255869>.

Boot, Armoud W.A., and Anjan V. Thakor. 1997. Banking scope and financial innovation. *Review of Financial Studies* 10, no. 4:1099–1131.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh. 2000. Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not). National Bureau of Economic Research Working Paper 7552, February.

Congressional Record. 1999a. Extension of Remarks in the House of Representatives, Thursday, August 5, 1999, American Inventors Protection Act of 1999, 145 *Cong. Rec.* E1788-02, at E1789 (Statement of Rep. Howard Coble, D.-N.C.).

———. 1999b. 145 *Cong. Rec.* S14,836 (daily ed. Nov. 18, 1999), at S.14,994 (statement of Sen. Schumer).

———. 1999c. House of Representatives, Thursday, November 18, 1999, Conference Report on H.R. 3194, Consolidated Appropriations and District of Columbia Appropriations Act, 2000 145 *Cong. Rec.* H12798-01, at H12805.

———. 1999d. 145 *Cong. Rec.* S14,836 (daily ed., Nov. 18, 1999), at S14,995 (statement of Sen. Torricelli).

Cowan, Robin, Paul David, and D. Foray. 2000. The explicit economics of knowledge codification and tacitness. *Industrial and Corporate Change* 9 (June): 211–53.

Davis, Julie L., and Suzanne S. Harrison. 2001. *Edison in the boardroom: How leading companies realize value from their intellectual assets*. New York: Wiley.

Frame, W. Scott, and Lawrence J. White. 2002. Empirical studies of financial innovation: Lots of talk, little action? Paper presented at the Federal Reserve Bank of Philadelphia conference, "Innovation in Financial Services and Payments," May 16–17.

Frankel, Tamar. 1998. Cross-border securitization: Without law, but not lawless. *Duke Journal of Comparative and International Law* 8 (Spring): 255–82.

Gans, Joshua Samuel, and Scott Stern. 2002. The product market and the market for 'ideas': Commercialization strategies for technology entrepreneurs. SSRN Electronic Library. <ssrn.com/abstract=317219>.

Hall, Bronwyn, and Rosemarie Ham-Ziedonis. 2001. The determinants of patenting in the U.S. semiconductor industry, 1980–1994. *Rand Journal of Economics* 32 (Spring): 101–28.

Heaton, J.B. 2000. Patent law and financial engineering. *Derivatives Quarterly* 7 (Winter): 7–15.

Hellmann, Thomas, and Manju Puri. 2000. The interaction between product market and financing strategy: The role of venture capital. *Review of Financial Studies* 13 (Winter): 959–84.

Merges, Robert P. 1999. As many as six impossible patents before breakfast: Property rights for business concepts and patent system reform. *Berkeley Technology Law Journal* 16 (Spring): 577.

———. 2000a. Intellectual property rights and the new institutional economics. *Vanderbilt Law Review* 53 (November): 1857–77.

———. 2000b. One hundred years of solicitude: Intellectual property law 1900–2000. *California Law Review* 88 (December): 2187–2240.

Merges, Robert P., and John F. Duffy. 2003. *Patent law and policy.* 3d ed. Newark, N.J.: Matthew Bender.

Merges, Robert P., and Richard R. Nelson. 1990. On the complex economics of patent scope. *Columbia Law Review* 90 (May): 839–916.

Persons, John C., and Vincent Warther. 1997. Boom and bust patterns in the adoption of financial innovations. *Review of Financial Studies* 10 (Winter): 939–67.

Polanyi, Michael. 1967. *The tacit dimension*. New York: Doubleday.

Red Herring. 2000. eSpeed is hungry for B2B markets. January 14.

Saloner, Garth, and Andrea Shepard. 1995. Adoption of technologies with network effects: An empirical examination of the adoption of automated teller machines. *Rand Journal of Economics* 26 (Autumn): 479–501.

Schmookler, Jacob. 1967. *Invention and economic growth*. Cambridge, Mass.: Harvard University Press.

Teece, David. 1986. Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy* 15, no. 6:285–305.

Tufano, Peter. 1989. Financial innovation and first-mover advantages. *Journal of Financial Economics* 25 (December): 213–40.

———. 2002. Financial innovation. Harvard Business School, unpublished paper. (Forthcoming in *The handbook of the economics of finance*, edited by George Constantinides, Milt Harris, and René Stulz. North Holland.)

Usselman, Steven W. 2002. *Regulating railroad innovation*. Cambridge: Cambridge University Press.

The International Law of Business Method Patents

JOHN M. CONLEY

The author is the William Rand Kenan Jr. Professor of Law at the University of North Carolina School of Law. This paper was presented at the Atlanta Fed's 2003 Financial Markets Conference, "Business Method Patents and Financial Services," cosponsored with the University of North Carolina School of Law.

In its 1998 decision in *State Street Bank and Trust Co. v. Signature Financial Group, Inc.*, the United States Court of Appeals for the Federal Circuit (which now hears all patent appeals in this country) addressed "the judicially-created, so-called 'business method' exception to statutory subject matter" (149 F. 3d 1368, 1375 [Fed. Cir. 1998], *cert. denied*, 525 U.S. 1093 [1999]). Throughout most of the history of American patent law, the courts and the U.S. Patent and Trademark Office (USPTO) had usually—but not uniformly—denied patents to inventions that amounted to nothing more than methods for doing business. In *State Street,* the Federal Circuit repudiated this long-standing practice in terms that could not have been blunter: "We take this opportunity to lay this ill-conceived exception to rest. . . . Since the 1952 Patent Act, business methods have been, and should have been, subject to the same legal requirements for patentability as applied to any other process or method" (*State Street*, 1375).

In the same decision, the Federal Circuit also repudiated the notion that computer-based inventions should be subject to special restrictions. Sweeping away three decades of complex and often inconsistent case law, the court held that a computerized process for transforming data is within the realm of patentable subject matter so long as it "produces a 'useful, concrete and tangible result'" (p. 1375). Whereas patent lawyers had previously felt it necessary to hide the computerized aspects of their patent claims in a conventionally patentable machine or process, *State Street* made it possible to bring software into the open.

Because contemporary business, particularly in the financial services area, is almost entirely dependent upon computers for its design and implementation, the interrelationship of the two *State Street* holdings is self-evident. Under previous law, it was widely believed that one could not patent either a pure business method or a pure software operation (that is, one that did not produce effects in the physical world). *State Street* allowed both, reversing the lower court's invalidation of a patent claiming the computerized implementation of a method of providing financial services. The broadest claim in the patent was drawn to "a data processing system for managing a financial services configuration of a portfolio established as a partnership, each partner being one of a plurality of funds," to be implemented by a generic system of hardware and software (p. 1371).

The *State Street* decision is perceived to have sparked a revolution in both law and business. One widely held view is that *State Street* made everything patentable in the business world and that business people are responding by trying to patent everything (Meurer, forthcoming). That may be something of an overstatement. Although business method patents were relatively uncommon before *State Street,* patent lawyers had found ways to obtain them and, on occasion, had successfully defended them in the courts (Kuester and Thompson 2001). Moreover, while *State Street* certainly led to an increase in the volume of business patent applications (Meurer, forthcoming), it has not been quite the flood that has been claimed. In addition, there is every possibility that here, as in other areas, what the Federal Circuit has given by expanding the standards for patentability it will take away by tightening the standards for enforcement.

Nonetheless, one cannot deny the extraordinary influence of the *State Street* decision, both legally and practically. If it did not quite revolutionize the law, it refined and restated it with absolute clarity. If nothing else, the publicity surrounding the *State Street* case in the legal and business worlds has created near-universal awareness of the existence and potential significance of business method patents.

This paper reviews the state of the law with respect to business method patents, both in the United States and internationally. It begins with a brief overview of the basic requirements for patentability in the United States and internationally. It presents in some detail the evolution and current state of American law and international law, focusing on the European Union, examples of European national law, and Japan. Finally, the paper analyzes legal trends both in the United States and abroad, makes concluding comparative comments, and offers some predictions about unfolding legal issues.

Basics of Patent Law

To meet the basic requirements for obtaining a patent under American law, an invention must pass four tests:

First, under Section 101 of the Patent Act of 1952 (35 U.S.C. §§ 100 *et seq.*), the patent application must claim so-called statutory subject matter. That is, it must claim a human-made process, machine, manufacture, or composition of matter, or an improvement thereon. Laws of nature, products of nature, and abstract ideas such as mathematical algorithms have historically been deemed nonstatutory (*Diamond v. Chakrabavty*, 447 U.S. 303 [1980]).

Second, the claimed invention must be novel. Novelty has a highly technical meaning, which is articulated in the complex provisions of Section 102 of the Patent Act. For example, under Section 102(a), the patent will be denied if the invention was known or used by others in this country, patented here or abroad, or described in a "printed publication" in the United States or a foreign country prior to the patent applicant's date of invention. Section 102(b) creates the "statutory bar" that results in a forfeiture of patent rights if the applicant or anyone else makes public use of the invention, puts it on sale, or engages in other specified conduct for more than a year prior to the filing of an application. Section 102(g) establishes the rules for determining priority when two or more inventors claim the same invention. American priority rules are virtually unique in international patent law: Priority is awarded to the person who can prove that he or she was the first to invent whereas in most other countries the patent goes to the first person to file a patent application.

The third requirement is utility. Although Section 101 requires that an invention be "useful," utility has no specific statutory definition, so its meaning is derived from case law. In the vast majority of instances, it is an easy standard to meet, requiring nothing more than a showing that the invention may be put to some beneficial (very broadly construed) use. Historically, chemistry has been the one area in which significant numbers of applications have been denied for lack of utility. In a 1966 case called *Brenner v. Manson* (383 U.S. 519 [1966]), for example, the Supreme Court denied a patent to "a chemical process which yields an already known product whose utility—other than as a possible object of scientific inquiry—has not yet been evidenced" (p. 532). The compound in question was closely related to a class of compounds that had been shown to inhibit tumors in mice—an unquestioned showing of utility—but whose own potential uses were not yet known. Following the same reasoning, the USPTO and the courts currently require that claims to genetic sequences disclose their function; it is not enough simply to state that the gene is an object of scientific inquiry that is ultimately likely to lead to beneficial medical applications.

The fourth and final requirement is nonobviousness. As set forth in Section 103(a) of the Patent Act, the specific rule is that the invention is unpatentable "if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which such subject matter pertains." The nonobviousness barrier will often trip up applicants who have survived the novelty inquiry. Under the novelty test, the patent will not be denied unless the very invention that is claimed has been described, used, etc. in its entirety before the critical date. Under the nonobviousness rule, by contrast, the patent will be denied if a hypothetical person of ordinary skill in the field, armed with the total knowledge in the field (the "prior art"), would have looked at the applicant's advance at the time it was made and deemed it an obvious step. As this description suggests, the nonobviousness requirement is highly subjective, and its application by the courts and the USPTO has been inconsistent over the years.

Assuming that these four standards can be satisfied, the application itself must meet certain formal requirements. The most important of these is Section 112's "enabling disclosure" rule. The patent application must describe the invention with enough specificity to enable a person skilled in the relevant field to make and use it. It is not necessary for the inventor actually to have built the invention (or in patent jargon, reduced it to practice) before filing the application. It is enough that the description provided in the application will enable someone else to build it and that the patent examiner is persuaded that it is indeed operable.

If a patent is granted, the inventor will be able to stop others from making, using, or selling the invention for the term of the patent (17 U.S.C. § 271[a]). In most cases, U.S. patents (as well as those in other countries) last for twenty years from the date the application is filed (35 U.S.C. § 154[a][2]). The words "make, use, and sell" are taken in their literal senses. The proscribed activities are strictly prohibited, regardless of whether they involve intentional copying or accidental duplication. The Patent Act also prohibits importing patented inventions into the United States from abroad (35 U.S.C. § 271[a]), as well as actively inducing others to commit acts of infringement (35 U.S.C. § 271[b]). Other provisions define a number of contributory infringements. These include knowingly selling or offering to sell specialized components of patented inventions (35 U.S.C. § 271[c]). U.S. patent laws, like most other national patent laws, generally lack extraterritorial effect, meaning that they do not cover most conduct outside the United States. However, it is also infringement to supply a specialized component of a patented invention from the United States, knowing that such component will be used abroad in a manner that would infringe the patent if done within the United States (35 U.S.C. § 271[f]) or to import a product of a patented process that is practiced abroad (35 U.S.C. § 271[f]). Successful patent infringement plaintiffs may be awarded injunctive relief, actual damages, and, in exceptional cases, multiple damages as well as attorneys fees (35 U.S.C. §§ 283–85).

Under a recent amendment to the Patent Act called the American Inventors Protection Act of 1999 (35 U.S.C. § 273), defendants accused of infringing business method patents have some special defenses. In general, it is a defense to an action for the infringement of a business method patent if the defendant, acting in good faith, had reduced the patented invention to practice (actually built it) more than one year before the plaintiff's application was filed and had used the invention commercially at any time before the plaintiff's filing. The defendant has the burden of proof to establish this defense and may not use it if he or she learned of the invention from the patent holder. Moreover, the defense is purely personal, and the defendant's right to use the invention may not be licensed or transferred to anyone else.

Whereas patent lawyers had previously felt it necessary to hide the computerized aspects of their patent claims in a conventionally patentable machine or process, *State Street* made it possible to bring software into the open.

The purpose of creating this new defense was to address a problem that is believed to be endemic in the business method patent area. At the time a business method application is being reviewed, the sources typically available to the patent examiner (principally, prior patents and conventional publications) may not reveal that the claimed invention was either not novel or obvious at the purported date of invention. Nonetheless, evidence may later emerge that others had been using the same technology well before the date of the application. For a variety of technical reasons, this prior use might not invalidate the patent. While these new provisions do not change the standards for patentability, they may prevent the patent holder from putting such prior users out of business.

A final point is that U.S. patent law is perhaps the most "back-end-loaded" in the world. The United States, in other words, is relatively lenient in granting patents, depending more heavily on judicial scrutiny when patentees bring infringement actions (Kesan 2002). Most other countries offer third parties a more meaningful opportunity to oppose a patent while it is pending or immediately after it is issued (Merges and Duffy 2002, 64). The U.S. law of reexamination has the effect of postponing most such challenges until the patentee brings an infringement action (35 U.S.C.

§§ 311–18).[1] Although plaintiffs' patents carry a presumption of validity, defendants can—and regularly do—attempt to show that patents were wrongly issued.

The substantive requirements for obtaining a patent vary little from country to country. For example, under Article 27 of the TRIPS Agreement (Agreement on Trade-Related Aspects of Intellectual Property Rights, enacted under the General Agreement on Tariffs and Trade), all members of the World Trade Organization are required to make patents available "for any inventions, whether products or processes, in all fields of technology, provided that they are new, involve an inventive step and are capable of industrial application." An accompanying footnote states, "The terms 'inventive step' and 'capable of industrial application' may be deemed by a Member to be synonymous with the terms 'nonobvious' and 'useful,' respectively." Similar standards have long been followed by Japan, the European Patent Office, and the individual member states of the European Union. As will be discussed later, there are material differences in patentability standards in some subject matter areas, including business methods and biotechnology.

U.S. Legal Doctrine

History. Despite the conventional view that patents on methods of doing business have long been disfavored, if not flatly prohibited, such patents have, in fact, been regularly granted. For example, the first financial services patent was probably granted to Jacob Perkins in 1789 for a system of detecting counterfeit notes; unfortunately, its details were lost in a fire in 1836 (USPTO 2000). In 1867 Charles L. Hawkes of Titusville, Pennsylvania, obtained a patent titled "Improvement in Hotel-Registers" (Letters Patent No. 63,889). His "invention" was to add to the margins of blank-ruled hotel register pages "advertisements of business houses, entertainments, railroad or steamboat cards, and other notices whose insertion is worth paying for." And in 1907 a patent was issued to Eugene Graves Adams of Lynchburg, Virginia, for an improved form for the accident insurance policies that were widely purchased by railway travelers of the age (Letters Patent No. 853,852). Adams claimed, "As an article of manufacture, a two-part insurance policy consisting of a paper containing an insurance contract . . . combined with a postal card, both bearing a number or mark of identification, to be mailed to the beneficiary."

Patents have regularly been granted on machines and processes intended to make business more efficient. In 1815, for example, John Kneas obtained a patent for an improvement in banknote printing (USPTO 2000). His advance was "to print copper plate on both sides of the note or bill, or copper plate on one side and letter press on the other side, or letter press on both sides of a bank note or bill as an additional security against counterfeiture." In 1889 Herman Hollerith obtained method and apparatus patents titled "Improvements in the Art and System of Computing Statistics" (Letters Patent No. 395,781). Hollerith's patents described the mechanical punch card system for processing business information that dominated the market until the age of personal computers. Hollerith founded the Tabulating Machine Company, whose name was changed to International Business Machines Corporation in 1924 by Thomas J. Watson Sr.

In spite of this history, the USPTO and most courts long recognized a nearly absolute prohibition against claims drawn to methods of doing business. The most often cited case is *Hotel Security Checking Co. v. Lorraine Co.,* a 1908 decision of the Second Circuit (160 F. 467 [2d Cir. 1908]). The patent in question involved a hotel bookkeeping system that provided for cash registering and account checking in a manner designed to prevent fraud. Although, as will be seen, the Federal Circuit in *State Street* treated *Hotel Security* as a case of novelty and nonobviousness rather than as a subject matter case, the Second Circuit did state that "a system of transacting business disconnected from the means for carrying out the system is not . . . an art" (p. 469). By "art," it meant "process" as that term is currently used in Section 101. This language was followed as settled law by a number of cases extending through the beginning of the computer age in the second half of the twentieth century. In a 1942 case called *In re Patton,* the Court of Customs and Patent Appeals (the Federal Circuit's predecessor) reaffirmed the *Hotel Security* doctrine by stating that a system for transacting business, separate from the means for carrying out that system, was not patentable subject matter (127 F. 2d 423 [C.C.P.A. 1942]). The USPTO followed the *Hotel Security* rule as well. Through 1996, Section 706.03(a) of the *Manual of Patenting Examining Procedures* contained the following statement: "Though seemingly within the category of process or method, a method of doing business can be rejected as not being within the statutory classes" (citing *Hotel Security Checking*).[2]

The seemingly absolute rule of *Hotel Security* began to erode in the 1960s and 1970s as computers were increasingly used to perform business functions. Claims drawn to computer-related inventions had a tortured history in the courts prior to *State*

Street. Two Supreme Court decisions may have contributed to the confusion. In its 1978 decision in *Parker v. Flook* (437 U.S. 584 [1978]), the Court rejected as nonstatutory a claim drawn to a method for calculating an "alarm limit" for catalytic converters that was intended to be implemented on a computer. The essential problem, as the Court saw it, was that the patent claimed nothing more than the calculation of a mathematical formula. Three years later, in *Diamond v. Diehr,* the Court upheld the statutory status of a claim on "a method of operating a rubber molding press for precision-molded compounds with the aid of a digital computer" (450 U.S. 175, 179 n.5 [1981]). The computer's function was the repetitive calculation of a well-known mathematical formula known as the Arrhenius equation. The Court apparently saw a material distinction between claiming an industrial process that happened to employ computer calculations and claiming the act of calculation itself as an aid to carrying out an industrial process.

Before and after the two Supreme Court decisions, the Court of Customs and Patent Appeals and its successor, the Federal Circuit, struggled with limited success to establish coherent rules for the patentability of computer-based inventions. Many cases focused on whether and under what circumstances the inevitable presence of mathematical algorithms in computerized processes would defeat the patent. Despite their inconsistency, these cases seemed to establish that the use of a computer to perform mathematical calculations would not in itself defeat patentability if the calculations were applied so as to affect or understand the physical world (*State Street,* 1373–75; Chisum 2002, § 1.03[6]). Accordingly, in 1992 the Federal Circuit upheld a patent claiming methods and apparatus for the computerized transformation of electrocardiograph signals into a form that would give a doctor useful diagnostic information (*Arrhythmia Research Technology, Inc. v. Corazonix Corp.*, 958 F.2d 1053 [Fed. Cir. 1992]).

Many patent lawyers drew a more straightforward lesson from a comparison of *Flook* and *Diehr*: A computer-based invention would survive statutory subject matter scrutiny so long as the functions of the computer were "hidden" in a familiar and otherwise patentable process or machine (Blumenthal and Riter 1980). Thus, even before *State Street,* patent drafters regularly obtained patents on processes that happened to include the operation of a computer or on machines that were nothing more than general-purpose computers programmed to perform the function in question (Merges and Duffy 2002, 151; Kuester and Thompson 2001; USPTO 2000). Means-plus-function claims were especially popular. In such claims, the function of a device is claimed and the general means for performing the function are recited; the specific structural features recited in the written description portion of the patent are then read back into the claims (35 U.S.C. § 112, ¶6). In its 1989 decision in *In re Iwahashi,* the Federal Circuit upheld a claim in this form on "an autocorrelation unit for providing autocorrelation coefficients for use as feature

The publicity surrounding the *State Street* case in the legal and business worlds has created near-universal awareness of the existence and potential significance of business method patents.

parameters in pattern recognition"—in other words, a device for implementing a mathematical algorithm for voice recognition purposes (888 F.2d 1370 [Fed. Cir. 1989]).

This growing tolerance of computer-based inventions spilled over into the business method area, leading to the allowance of a number of patents on methods of doing business that were implemented by computerized means (Chisum 2002, § 1.03[5]). In 1974, in *In re Johnston* (502 F.2d 765 [C.C.P.A. 1974]), the Court of Customs and Patent Appeals found that a patent drawn to an automatic record-keeping system for a bank constituted statutory subject matter. Perhaps significantly, the claim was on a machine—a digital computer programmed to operate the system—rather than on the process itself. Nine years later, in *Paine, Webber, Jackson and Curtis, Inc. v. Merrill, Lynch, Pierce, Fenner & Smith, Inc.* (564 F. Supp. 1358 [D. Del. 1983]), a federal district court in Delaware rejected a subject matter challenge to a claim on a "securities brokerage-cash management system." The relevant claims, drafted in means-plus-function form, were directed to

1. A third party who requests reexamination and loses may not challenge validity of the patent in subsequent infringement litigation "on any ground which the third-party requester raised or could have raised" in the reexamination (35 U.S.C. § 313[c]).
2. The manual and all other official publications of the USPTO are available on-line at its Web site, <www.uspto.gov>.

computer hardware and software, designed and programmed to implement a system whereby the brokerage could manage all aspects of customer accounts. Paine Webber, seeking a declaratory judgment of noninfringement, attacked the patent as claiming "nothing more than familiar business systems, that is, the financial management of individual brokerage accounts" (p. 1365). Citing prior decisions of the Court of Customs and Patent Appeals, the district court held that "the product of a computer program is irrelevant, and the focus of analysis should be on the operation of the program on the computer" (p. 1369). Therefore, it concluded, the Merrill Lynch patent passed the statutory subject matter test as "a method of operation on a computer to effectuate a business activity" (p. 1369).

U.S. patent law is perhaps the most "back-end-loaded" in the world. The United States is relatively lenient in granting patents, depending more heavily on judicial scrutiny when patentees bring infringement actions.

The USPTO's Board of Patent Appeals and Interferences stated the evolving doctrine succinctly in its 1988 decision in *Ex Parte Murray:* "Whereas an apparatus or system capable of performing a business function may comprise patentable subject matter, a method of doing business generated by the apparatus or system is not" (9 U.S.P.Q. 2d 1819, 1820 [Bd. Pat. App. & Interf. 1988]). *Murray* held that the "claimed accounting method, requiring no more than the entering, sorting, debiting, and totaling of expenditures as necessary preliminary steps to issuing an expense analysis statement, is, on its very face, a vivid example of the type of 'method of doing business' contemplated by our review court [the Federal Circuit] as outside the protection of the patent statutes" (p. 1820).

The distinction drawn by the board in *Murray* is useful in explaining other post-computer but pre-*State Street* business method cases. For example, in *In re Maucorps* (609 F. 2d 481 [C.C.P.A. 1979]) and *In re Meyer* (688 F.2d 789 [C.C.P.A. 1982]), the Court of Customs and Patent Appeals rejected as nonstatutory claims drawn, respectively, to a business methodology for deciding how salesmen should best handle particular customers and a system for aiding neurologists in diagnosing patients. Then, in the 1994 case of *In re Schrader,* the board denied statutory status to a claimed system of auction bidding and the Federal Circuit affirmed (22 F. 3d 290 [Fed. Cir. 1994]). While the board relied both on the abstract mathematical algorithm and the business method exceptions, the Federal Circuit's majority opinion focused only on the former. In a significant dissent, Judge Pauline Newman took the opportunity to review the history of the business method doctrine and concluded that it "merits retirement from the glossary of Section 101" (pp. 296–98). She distinguished a number of often-cited business method cases (including *Hotel Security*) as being better analyzed as novelty or nonobviousness cases. She argued that "historical distinctions between a method of 'doing' business and the means of carrying it out blur in the complexity of modern business systems" (p. 298), thus rejecting the analysis suggested by *Murray*. She also quoted the Delaware district court's *Merrill Lynch* opinion approvingly and at length. As will be seen in the next section, Judge Newman's conclusion and reasoning were to be adopted almost unchanged in *State Street.*

A final development was the USPTO's deletion of the business method prohibition from the *Manual of Patent Examining Procedures* in 1996. Simultaneously, the following language was added to the 1996 edition of the *Examination Guidelines for Computer-Related Inventions*: "Office personnel have had difficulty in properly treating claims directed to methods of doing business. Claims should not be categorized as methods of doing business. Instead, such claims should be treated like any other process claims" (61 Fed. Reg. 7478, 7479 [1996]). The USPTO's more explicitly flexible attitude was quickly reflected in its examination results. The late 1990s saw the issuance of a significant number of patents on what appeared to be standard business practices conducted on the Internet (Oxford IPRC 2000, 17–18; Meurer, forthcoming, 6).

The state of American law with respect to business method patents immediately prior to the *State Street* decision can be summarized as follows: To the extent there had ever been an absolute bar on patenting methods of doing business, it had all but disappeared. Filings in the USPTO were becoming more numerous and more aggressive. The USPTO itself had moved from intransigence to flexibility to what some regarded as abject surrender in the face of such filings. The courts, meanwhile, were not always consistent but were, on balance, increasingly accommodating. Drawing on the proliferating case law concerning computer-based inventions, some courts attempted to draw distinctions between claims to pure business methods, which remained nonstatutory, and claims to otherwise patentable machines and systems (programmed computers, in both cases, whose purpose was to implement business methods).

One final point to be emphasized is that these legal controversies focused on the question of statutory subject matter status. The novelty, utility, and nonobviousness inquiries were always (at least in theory) conducted in exactly the same way as they were with respect to any other kind of invention.

The *State Street* and *AT&T* Decisions. The *State Street* case involved a patent (U.S. Pat. No. 5,193,056) that had been issued to Signature Financial Group, Inc., in 1993, titled "Data Processing System for Hub and Spoke Financial Services Configuration" (p. 1370). State Street, like Signature, is in the business "of acting as custodians and accounting agents for multi-tiered partnership fund financial services" (p. 1370). When State Street was unable to negotiate a license to use the Signature patent, it filed suit, seeking a declaratory judgment of invalidity and noninfringement. The Massachusetts district court granted summary judgment for State Street on the issue of invalidity, and the Federal Circuit ultimately reversed.

More specifically, the patented invention allows for the unified management of a portfolio set up as a partnership, with each partner being a separate mutual fund. The portfolio is characterized as the "hub" and the constituent funds as "spokes." The "system provides means for a daily allocation of assets for two or more Spokes that are invested in the same Hub" (p. 1371). It "determines the percentage share that each Spoke maintains in the Hub, while taking into consideration daily changes both in the value of the Hub's investment securities and in the concomitant amount of each Spoke's assets" (p. 1371). The system allocates the hub's daily income, expenses, and net realized and unrealized gains or losses among the constituent spokes. This allocation allows for the calculation of the true asset value of each spoke on a daily basis as well as for the year-end aggregation of income, expenses, and capital gain or loss. Because each spoke is a mutual fund selling shares to the public, it is essential for pricing purposes that it has real-time data based on its percentage interest in the hub portfolio.

Signature's application, filed in 1991, initially contained six machine claims in means-plus-function form as well as six method claims. Signature cancelled the method claims in response to the patent examiner's opposition, and the six means-plus-function claims were ultimately allowed. The only independent claim,[3] claim 1, recited "a data processing system for managing a financial services configuration of a portfolio established as a partnership, each partner being one of a plurality of funds, comprising" a variety of computer hardware and software means (p. 1371). The district court treated this and the other five claims as process claims and rejected them because the mathematical algorithm that they included was not "applied to or limited by physical elements or process steps" (927 F. Supp. 508, 513 [D. Mass. 1996]). Drawing on the pre-*State Street* case law, the district court concluded, not unreasonably, that the patent claimed an abstract mathematical calculation that was not adequately tied to the physical world. The district court also observed that its decision "comports with another doctrinal exclusion from subject matter patentability known as the 'business methods exception'" (p. 515). It cited numerous treatises and cases for the continuing validity of the doctrine and, in particular, for the developing distinction between an apparatus or a system capable of performing a business function and that function itself.

The Federal Circuit thoroughly repudiated both aspects of the district court's decision. Initially, it observed that the claims were properly viewed as being in machine rather than process form, although the distinction would ultimately prove immaterial. It then significantly narrowed the mathematical algorithm exception to patentability, thereby clarifying and simplifying the law of computer-related patents. Citing the Supreme Court's 1981 decision in *Diamond v. Diehr* (discussed above), the Federal Circuit acknowledged "that mathematical algorithms are not patentable subject matter to the extent that they are merely abstract ideas" (p. 1373). The court went on, however, to redefine radically what is meant by "abstract." Specifically, "to be patentable, an algorithm must be applied in a 'useful' way" (p. 1373). On the facts before it, the court held "that the transformation of data, representing discrete dollar amounts, by a machine through a series of mathematical calculations into a final share price,

3. An independent claim, as the word suggests, stands alone and is interpreted without reference to any others. A dependent claim incorporates the claim on which it depends and then adds further limitations. An independent claim might, for example, recite a chemical process, and a subsequent dependent claim could incorporate the first claim but then require that it be carried out in a specified pH range.

constitutes a practical application of a mathematical algorithm, formula or calculation, because it produces a 'useful, concrete, and tangible' result—a final share price momentarily fixed for recording and reporting purposes" (p. 1373). It repudiated a prior test (the so-called *Freeman-Walter-Abele* test) that focused on the application of algorithms to physical elements as having "little, if any, applicability to determining the presence of statutory subject matter" (p. 1374). Henceforth, the only test is whether computation of the algorithm yields a useful, concrete, and tangible result. A dollar number that will be of use in the financial services industry constitutes such a result.

***State Street* greatly simplified the law.... A computer-based invention now constitutes patentable subject matter so long as the computer operation produces a specific and useful result even if that result is simply a number.**

The court then turned to the business method exception and disposed of it succinctly and summarily: "We take this opportunity to lay this ill-conceived exception to rest" (p. 1375). The exception should not have survived the 1952 Patent Act's all-inclusive definition of statutory subject matter. Tracking Judge Newman's dissent four years earlier in *Schrader,* the court expressed doubt that the doctrine had ever been as robust as generally assumed. Analyzing its own precedent, as well as that of the Court of Customs and Patent Appeals, the court noted that "[a]pplication of this particular exception has always been preceded by a ruling based on some clearer concept" (p. 1375). Even *Hotel Security*, the court found, again following Judge Newman, was really a novelty and nonobviousness case. Finally, the court also endorsed the proposition that the purported distinction between a method of doing business per se and the means of implementing that method was far too fuzzy to be of any ongoing utility. Therefore, the court concluded, "Whether the claims are directed to subject matter within Section 101 should not turn on whether the claimed subject matter does 'business' instead of something else" (p. 1377).

In summary, in a factual context that reflects the inseparability of computer technology and modern financial services, the Federal Circuit significantly enhanced the patentability of both business methods and computer-based inventions generally. With respect to the former, it dismissed as irrelevant the characterization of a patent claim as drawn to a method of doing business. With respect to the latter, it cut through a convoluted case law to hold that computer systems implementing mathematical algorithms can constitute statutory subject matter so long as they produce a useful, concrete, and tangible result. Finally, and perhaps most significantly, it bridged the two legal points by holding that a set of numbers of use to the financial services community constitutes precisely such a result.

A year later, the Federal Circuit decided *AT&T Corp. v. Excel Communications, Inc.* (172 F.3d 1352 [Fed. Cir. 1999]). The invention in that case involved a system for creating message records for long-distance telephone calls. Whereas the patent in *State Street* was characterized as a means-plus-function machine claim, AT&T's patent contained ten method claims. A Delaware federal district court had held the patent invalid under Section 101 for want of statutory subject matter. The Federal Circuit reversed that decision.

AT&T did not directly involve the business method exception, but it is relevant to the topic as a reaffirmation of *State Street*'s impact on computer-based inventions. The district court recognized that the claimed method required the use of computers and switches. The court held the method nonstatutory, however, on the grounds that it involved nothing more than the operation of a mathematical algorithm without any physical steps. The algorithm in question was basic Boolean algebra.

In reversing the decision, the Federal Circuit focused on the variable being calculated, "the PIC indicator value." The PIC indicator value provides a record of a customer's primary long-distance service carrier. Therefore, in the telephone business it is "a useful, non-abstract result that facilitates differential billing of long-distance calls" (p. 1358). Just like the financial data produced by the system in *State Street,* the production of a PIC indicator value was a sufficiently useful, concrete, and tangible application of the Boolean algorithm as to "fall comfortably within the broad scope of patentable subject matter under Section 101" (p. 1361).

AT&T v. Excel put to rest any concern that the *State Street* court did not mean what it said about the patentability of computer-based inventions. Once again, a process that does not produce an effect in the physical world has been held nonetheless to be "useful, concrete and tangible." In other words, "tangible" really means "specific." Combined with the demise of the business method barrier, this holding means that any computer-based invention that performs a business or financial operation should be patentable subject matter. In every case, of course, the other standards of patentability—novelty, utility, and nonobviousness—will still have to be satisfied. *AT&T* illustrates this latter point: on remand to consider these other factors, the district court invalidated the patent on novelty and obviousness grounds (1999 U.S. Dist. LEXIS 17871 [D. Del. 1999]).

Subsequent Developments. The state of the law can be categorized as stable. Perhaps the most closely watched case has been *Amazon.com v. barnesandnoble.com* (239 F.3d 1343 [Fed. Cir. 2001]). Amazon.com sued for infringement of its patent on a "method and system for placing a purchase order via a communications network" (U.S. Pat. No. 5,960,411). The claims, which were drafted with great breadth, cover one-click on-line shopping, both with and without the use of a shopping basket. Amazon filed the suit in its hometown district court in the state of Washington and was granted a preliminary injunction on December 1, 1999. The award of a preliminary injunction requires a finding that the plaintiff has a probability of success on the merits. The district court was therefore required to find that Amazon would probably succeed on the issue of patent validity. The Federal Circuit vacated this injunction in 2001, expressing doubts about Amazon's ability to defend the validity of the patent. Significantly, these doubts arose under Sections 102 and 103—not 101. Therefore, one should not read into this decision any doubts about the *State Street* and *AT&T* decisions. The case was settled on undisclosed terms before the district court rendered a final decision on the merits (Merges and Duffy 2002, 1052).

Summary. The *State Street* case has officially killed off whatever was left of the outright subject matter ban on patenting methods of doing business. Indeed, such patents are no longer even in the disfavored category. Simultaneously, *State Street* greatly simplified the law with respect to computer-based inventions. A computer-based invention now constitutes patentable subject matter so long as the computer operation produces a specific and useful result even if that result is simply in the form of a number.

State of International Legal Doctrine

Europe. This section will deal with two topics: legal developments concerning business method patents in Europe as a whole and related developments in individual European countries.

Business methods and "European" patents. The first and perhaps most significant point to be made is that there is at present no such thing as a true European patent (Merges and Duffy 2002, 55–56; Taketa 2002, 962–64). There are currently three ways to obtain a patent in Europe: proceeding through (1) the European Patent Office in Munich, (2) individual national patent offices, and (3) the Patent Cooperation Treaty. Since the Patent Cooperation Treaty is a procedural agreement intended primarily to assist countries with limited resources in processing applications, it will not be discussed further here.

The European Patent Office (EPO) was established in 1973 under the European Patent Convention (EPC).[4] The EPO is a hybrid organization with both procedural and substantive functions. Although all European Union members are signatories to the EPC, the EPO is an intergovernmental rather than EU body. An applicant files a single application with the EPO, designating the particular EPC countries in which patent protection is sought. The EPO then conducts a single examination of the application under unitary patentability standards established by the EPC. What is issued, however, is not a true European patent but a bundle of national patents. (An ongoing EU effort to develop a unitary European patent is discussed below.) Significantly, a patent holder is required to file infringement actions in the national courts of the countries in which infringement is alleged. This requirement is, of course, expensive and inefficient—in contrast to the situation of a U.S. patent holder whose single federal patent, enforceable in the federal courts, covers the entire United States. Moreover, although the enforcing European courts theoretically apply the same law, there is a substantial risk of variable interpretations. Again, this situation is in contrast with that in the United States, where all patent appeals go to the Federal Circuit.

The general EPC standards for patentable subject matter do not differ substantially from their American counterparts. Under Article 52(1) of the EPC, "European patents shall be granted for any inventions which are susceptible of industrial application, which are new and which involve an inventive step." These three requirements are generally viewed as equivalent to the American criteria of utility, novelty, and nonobviousness.

4. For general information on the EPO and the EPC, see the EPO's Web site at <www.european-patent-office.org/epo-general.htm>. The text of the EPC is available at the same site at <www.european-patent-office.org/legal/epc/index.html>.

The specific standards governing business method patents show similarities and differences when compared to the American rules. Article 52(2) contains a number of specific exclusions, including "schemes, rules and methods for performing mental acts, playing games or doing business, and programs for computers." This apparently explicit prohibition against patenting either computer programs or methods for doing business is not nearly so absolute as it appears, however. The next section, Article 52(3), states that "the provisions of paragraph 2 shall exclude patentability of the subject matter or activities referred to in that provision only to the extent to which a European patent application or European patent relates to such subject matter or activities *as such*" (emphasis supplied).

The holy grail of "technical character" seems little more than a challenge to European claim drafters.... Business methods will be found to be patentable subject matter, if not through the front door then through the back.

According to an official EPO press release on business methods and computer programs, the phrase "as such" is critical:

> It follows that, although methods for doing business, programs for computers, etc. are *as such* explicitly excluded from patentability, a product or a method which is of a technical character may be patentable, even if the claimed subject matter defines or at least involves a business method, a computer program, etc. (EPO 2000)

The recent EU Commission document proposing a directive on computer-implemented inventions (discussed below) makes two related points (Comm. of the EC 2002, 7–8). First, "an algorithm which is considered as a theoretical entity in isolation from the context of a physical environment, and in respect of which it is accordingly not possible to infer its effects, will be inherently non-technical and thus not susceptible of being regarded as a patentable invention." However, the second point—"all programs when run in a computer are by definition technical"—virtually moots the first. An algorithm apparently becomes "technical," and thus potentially patentable, so long as it is implemented on a computer.

The current *Guidelines for Examination in the EPO* reinforce these principles.[5] Under the heading "Schemes, Rules and Methods for Performing Mental Acts, Playing Games, or Doing Business," the guidelines state:

> These are further examples of items of an abstract or intellectual character. In particular, . . . a scheme for organizing a commercial operation would not be patentable. However, if the claimed subject matter specifies an apparatus or technical process for carrying out at least some part of the scheme, that scheme and the apparatus or process have to be examined as a whole. In particular, if the claim specifies computers, computer networks or other conventional programmable apparatus, or a program therefor, for carrying out at least some steps of a scheme, it is to be examined as a "computer-implemented invention."

The next section, "Programs for Computers," summarizes the relevant doctrine as follows:

> When considering whether a claimed computer-implemented invention is patentable, the following is to be borne in mind. In the case of a method, specifying technical means for a purely nontechnical purpose and/or for processing purely nontechnical information does not necessarily confer technical character on any such individual step of or use on the method as a whole. On the other hand, a computer system suitably programmed for use in a particular field, even if that is, for example, the field of business and economy, has the character of a concrete apparatus, in the sense of a physical entity or product, and thus is an invention within the meaning of Article 52(1).
>
> The same section states elsewhere:
>
> [Computer-implemented invention] claims may, e.g., take the form of a method of operating said conventional apparatus, the apparatus set up to execute the method, or following [a decision of the EPO Boards of Appeal], the program itself. Insofar as the scheme for examination is concerned, no distinctions are made on the basis of the overall purpose of the invention, i.e., whether it is intended to fill a business niche, to provide some new entertainment, etc.

It is difficult to distinguish these principles in material ways from the current state of U.S. law. First, as *State Street* did, the EPO *Guidelines* make it clear that a claim directed to carrying out a business method is not for that reason barred or even disfavored. Second, claims on computer-implemented inventions generally are also neither barred nor disfavored. Third, such claims may be drafted either in apparatus (machine) or in process form, specifying computers, computer networks, or even software (recall that *State Street* involved a machine claim, whereas the patent in *AT&T* claimed a method or a process). Fourth, an algorithm "as a theoretical entity in isolation" is not a patentable invention, a rule not materially different from *State Street*'s requirement that a claimed algorithm must be tied to a useful, concrete, and tangible result. In each of these respects, the EPO position seems wholly consistent with the doctrine of *State Street* and *AT&T.* To the extent that there is a difference, it is that the EPO, with its "as such" doctrine, is insisting on the distinction, drawn in the now-repudiated *Murray* case, between "an apparatus or system capable of performing a business function . . . [and] a method of doing business generated by the apparatus or system."

A fifth aspect of the EPO subject matter requirements, the so-called technicality or technicity standard, is ostensibly distinguishable from the U.S. standards but is likely to yield functionally similar results in many cases (EU 2001). Under the EPO *Guidelines* for "Programs for Computers," the claimed invention must have "technical character." This requirement is satisfied if "technical considerations are required to carry out the invention," and such technical considerations must be reflected in the claims. A technical consideration will be found, however, in the case of "a computer system suitably programmed for use in a particular field, even if that is, for example, the field of business and economy." This is to be contrasted with "a method, specifying technical means for a purely nontechnical purpose," which would not be patentable. Putting these various principles together, it appears that the technical character requirement will be satisfied by any computer, computer network, or computer program that is developed or improved to yield a specific result in a particular practical field of endeavor.

Although *State Street* and *AT&T* do not contain similar language, they achieve a similar effect. Their principal holdings are (1) that the mathematical algorithms embodied in computer programs do not bar patentability so long as their use produces a useful, concrete, and tangible result and (2) that the production of specific business or financial data satisfies that criterion. Thus, although worded differently, the U.S. and EPO subject matter standards seem to be functionally similar.

The EPO *Guidelines* are derived from the case law of the EPO Boards of Appeal. Perhaps the most important of its business method decisions is the *Sohei* case (T 769/92, 1995 OJ EPO 525 [1994]),[6] which is cited in the *Guidelines* and has been widely discussed in the European literature (Oxford IPRC 2000, 35). In *Sohei,* the applicant claimed "a computer system for plural types of independent management including at least financial and inventory management" and a method for operating said system. Data could be input using a single "transfer slip," which could take the form of an image displayed on a computer screen. The board held that the claimed subject matter constituted an invention under Article 52(1) of the EPC and could not be excluded from patentability under Articles 52(2)(c) and (3).

Consistent with the U.S. practice of treating patentable subject matter as an initial inquiry independent of novelty and nonobviousness, the appellant Sohei argued that "technicality . . . of an invention should, in principle, be examined independently of the question of novelty and inventive step." The board apparently agreed, "remitting" the case to the EPO's Examining Division for further consideration of the questions of novelty and inventive step. Sohei then argued that a computerized invention such as that claimed could not be held unpatentable under Article 52 as a program for a computer "as such":

> Whenever a computerized solution of a problem involves an implementation which is different from how a human being would solve the problem manually or mentally, technicality in the above sense should be assumed. As to computer programs, Article 52(2)(c) was only intended to exclude program listings.

Although the board did not endorse so broad a proposition, it did find in Sohei's favor. The claimed invention embodied adequate technicality because "the file handling needs a knowledge of the capacities of the computer on which the respective program is to be run." The claim in question was really directed to the operation of the computer system, which is technical; the financial and inventory management systems, which are not technical, were held to be

5. The guidelines are available at the EPO's Web site, <www.european-patent-office.org/legal/gui_lines/e/c_iv_2.htm>.
6. The EPO Boards of Appeal decisions are available at the EPO Web site at <legal.European-patent-office.org/dg3/search_dg3.htm>.

tangible illustrations of the operation of the system and perhaps "a voluntary limitation of the scope of protection." Their mention, however, did not undercut the technicality of the invention. Finally, the fact that the computer system was to be used to implement business methods that might be unpatentable "as such" did not render the system itself unpatentable: "Against claims so generalized in the Board's view, no objection that they relate only to 'doing business' as such could be raised." As the Oxford Intellectual Property Research Centre has put it, "the Board attached no importance to the end use of the system" (2000, 35). The bottom line is that, despite apparently more complex and demanding requirements, a computerized system for solving a pure business problem constitutes patentable subject matter under the European Patent Convention almost to the same extent as under the *State Street* regime.

Japanese law and practice with respect to both software and business method patents are usually described as being similar to the EPO's, with both standing in contrast to the American situation.

A more recent board decision underlines the importance of claim drafting to the determination of whether a business method constitutes patentable subject matter. The *Pension Benefits Systems Partnership* (TT 931/95 [2000]) case involved two primary claims: the first drawn to "a method of controlling a pension benefits program" that involves various unspecified "data processing means" and "computing means" and the second claiming "an apparatus for controlling a pension benefits system" that involves "data processing means." (Interestingly, the applicant is an American company.) The board held that the first claim "does not go beyond a method of doing business as such, and therefore, is excluded from patentability under Article 52(2)(c) in combination with Article 52(3) EPC; the claim does not define an invention within the meaning of Article 52(1)." The board rejected the argument that the references to data processing and computing means "conferred technical character to the method claimed," finding instead that the method amounted "to no more than the general teaching to use data processing means for processing or providing information of purely administrative, actuarial and/or financial character." The apparatus claim, however, was upheld as "constituting a physical entity or concrete product suitable for performing or supporting an economic activity."

These sorts of distinctions are insubstantial, if not illusory. The doctrine that emerges resembles the muddle that characterized U.S. case law before *State Street.* The holy grail of "technical character" seems little more than a challenge to European claim drafters.[7] There is no reason to doubt that, like their American counterparts, they will be up to it. Business methods will be found to be patentable subject matter, if not through the front door then through the back.

A more substantial distinction appears to lie in the EPO's application of the inventive step (nonobviousness) requirement to business method and computer-related inventions. The EPO requires that the inventive step be in a technical area; thus, an obvious computer implementation of a nonobvious business method will fail. In a consultation paper prepared to guide discussion on the proposed Directive on Computer-Implemented Inventions, the EU technical staff emphasized that "[t]he fact that the technical contribution also has to be non-obvious is an important limitation on the patentability of computer-implemented inventions" (Comm. of the EC 2000, 4). *State Street* does not appear to contemplate such a limitation.

Nonetheless, it is not clear that EPO examination practice is significantly more onerous than that in the USPTO. The examination process begins with the presumption that business methods are not per se unpatentable. The examiner next looks for an inventive step; to satisfy this criterion the invention must solve a technical problem. However, "if implementation of a business method calls for solution of a technical problem, it will pass muster"; "the overall purpose of the invention is not considered material" (Oxford IPRC 2000, 36). A conventional novelty inquiry follows.

With respect to software-based inventions generally, the president of the EPO stated in 1998 that, "Far from being antisoftware, we have been at pains to ensure that the European Patent system remains fully in tune with the needs of the software industry. . . . The EPO's approach to software-related inventions has been liberal" (Oxford IPRC 2000, 39). The ultimate question is whether this liberality will extend to software-based business method inventions. Perhaps spurred by the *Sohei* decision, EPO business method applications have risen substantially in the last few years (the vast majority are still pending) although the volume is as yet nowhere near what it has been in the United States (pp. 40–41). This lag may be due to real differences in legal requirements for patenting. It is equally plausible, however, that the disparity results from late-developing awareness among European companies of the availability of business method patents. Indeed, a recent Oxford IPRC survey indicates that most EPO business method filings are made by U.S. nationals (p. 41, app. B). According to the same survey, it remains too early to discern whether the EPO will distinguish itself from the USPTO in the handling of business method applications.

European national laws. The national law with respect to computer-implemented inventions in general, and business method inventions in particular, is well developed only in the United Kingdom and Germany. U.K. courts deal with the interpretation and enforcement of U.K. patents issued by both the U.K. Patent Office (UKPO) and the EPO. Although the U.K. courts are not bound by the decisions of the EPO Boards of Appeal, they are influenced by a parliamentary declaration of intention that patent laws be uniform throughout the EU. Nonetheless, the recent Oxford IPRC report concludes that "United Kingdom courts approach the issue of excluded subject matter in a manner somewhat less favorable to the patentee [than the EPO]" (2000, 37).

The critical difference may be that, whereas under EPO law the ultimate objective is irrelevant to the patentability of a computer system, a 1996 English decision rejecting a patent on a program for designing chemical structures held that "the Court or Patent Office must direct its attention not to the fact that the program is controlling the computer but to what the computer, so controlled, is doing" (*Fujitsu Ltd.'s Application* [1996] RPC 511 [Pat. Ct.], *aff'd* [1997] RPC 561 [Ct. App.]). This view is consistent with the 1989 English Court of Appeal decision in *Merrill Lynch Inc.'s Application* ([1989] RPC 561 [Ct. App.]), which held unpatentable a data processing system for buying and selling securities. In contrast to the approach taken by the EPO in *Sohei,* the *Merrill Lynch* court held that, although a data processing system operating to produce a novel technical result would normally be patentable, such a system is unpatentable "if the result itself is a prohibited item" such as a method of doing business. The UKPO (2001) has recently reaffirmed its adherence to these principles and its intent to do so for the foreseeable future.

German case law, by contrast, has been interpreted as "not exclud[ing] the possibility that business methods having a technical aspect could be patentable, even if the only contribution that the invention makes is nontechnical" (Comm. of the EC 2002, 10). A recent decision of the German Supreme Court has emphasized that German courts should follow the EPO approach and require that the inventive step constitute a technical contribution (p. 10).

Japan. Japanese law and practice with respect to both software and business method patents are usually described as being similar to the EPO's, with both standing in contrast to the American situation. For example, the background material to the EU's proposed directive on computer-implemented inventions states that "in Europe there has to be a *technical contribution* provided by the invention. In Japan there is a doctrine which has traditionally been interpreted in a similar way: the invention has to be a highly advanced creation of technical ideas by which a law of nature is utilized" (Comm. of the EC 2002, 5).

Japan has no outright ban on either software or business method patents. On the contrary, with respect to business methods the stated policy of the Japanese Patent Office (JPO) is "to offer appropriate protection of intellectual property rights (IPRs) in this field under close cooperation with overseas national patent offices" (JPO 2000). The JPO examines business methods applications under the category of "Computer Software-Related Inventions." It issued highly detailed new *Examination Guidelines* for such inventions on December 28, 2000.[8] These new guidelines suggest that business method claims (at least those that are implemented by computers) may pass the statutory subject matter test almost as easily as they do in the United States but that the inventive step scrutiny will approximate that in the EPO.

According to the new JPO guidelines, business method claims face three major hurdles: statutory subject matter, the requirement that inventions be "clearly stated," and inventive step. To meet the subject matter requirement of "a creation of technical ideas utilizing a law of nature," a business method

7. Two recent EPO board decisions concerning software, both captioned *International Business Machines Corporation,* have further complicated this issue (T 1173/97 [1998] and T935/97 [1999]). Both involved claims drawn to "a computer program product." In both, after long and convoluted discussion of the nature of technical character, the board remitted the application back to the examiners to continue the search for this elusive prey. Both patents ultimately issued; there were some modifications, but both still contained computer program product claims.
8. These guidelines are available at the JPO's Web site at <www.jpo.go.jp/tetuzuki_e/index.htm>.

or other software-related invention must be "concretely realized by using hardware resources" (p. 11). The JPO's examples of business method claims that meet this standard include "a storing method of articles distributed via network" (p. 33), "a computer program for predicting daily sales of commodities" (p. 36), and "a service method for offering service points depending on an amount of commodity purchased in telephone shopping" (p. 43). These examples suggest that it makes no difference whether a business method claim is drawn to the method of providing the service or the implementing software.

The requirement that an invention be "clearly stated" is—ironically if not unexpectedly—not very clear. The following is an example of a claim that comes up short: "an order-receiving method using a computer, comprising the steps of . . ." (p. 4). The problem with this claim is that it is unclear whether it is to be construed "as an order-receiving method (by a human) using a computer as a mere calculation tool" or "as an information processing method by computer software in the constructed order-receiving system" (p. 5). A claim to "a program equipped with an order-receiving means to accept a commodity order from a customer" is said to be similarly flawed but easily curable by amendment to "a program to make the computer operate as an order-receiving means" (p. 5). On balance, though this issue receives substantial attention in the new guidelines, it seems to be little more than a technical challenge to Japanese patent lawyers.

The substantial rhetorical gap between the United States on the one hand and Europe and Japan on the other concerning the patentability of business methods might not be so profound in practice.

The inventive step question is far more substantive. The basic concept is very much like the U.S. nonobviousness standard: whether "a person skilled in the art could easily have arrived at a claimed invention based on cited inventions" (p. 15). As in the United States, the claimed invention is to be viewed as a whole. The JPO guidelines then offer extended examples of inventions that will fail the inventive step test; two categories are especially relevant to business methods. The first is the application of existing knowledge to other fields. For example, "[w]here there exists the cited invention of 'medical information retrieval system', to apply the concrete means for retrieving in said 'medical information retrieval system' to a 'commodity information retrieval system' is deemed to be within the ordinary creative activity of a person skilled in the art" (p. 16). This example appears to involve the same general category of invention as the EPO's *Sohei* decision, where the Boards of Appeal found patentable subject matter but left open the question of inventive step.

The second noninventive category is the "systematization of human transactions," in which "the cited prior art describes human transactions but not how to systematize them" (p. 17). Business examples include "[m]erely to replace a telephone or fax previously used in order to receive orders from customers with a home page on the Internet," and "[m]erely to change the way of managing a classified section in a magazine into a way of managing such information via the home page on the Internet" (p. 17). These examples are reminiscent of the patent in dispute in the *Amazon.com v. barnesandnoble.com* case in the United States. The USPTO had issued the patent, but the Federal Circuit was dubious whether it would hold up under novelty and nonobviousness scrutiny. The overall import of these inventive step examples seems to be that the JPO will unequivocally oppose patents that lie in a gray area in the United States and maybe even in the EPO.

Whatever the theoretical distinctions, there is evidence that the JPO's results do not differ materially from those reached by the USPTO. At a meeting in Japan in the summer of 2000, the "Trilateral Offices" (the JPO, the USPTO, and the EPO) carried out an interesting experiment (Trilateral Technical Meeting 2000). The JPO and the USPTO examined several sets of hypothetical business method claims. Despite some differences in their respective approaches to statutory subject matter, the two offices resolved the issues of novelty and inventive step in virtually identical fashion and, consequently, arrived at the same results on the ultimate issue of patentability. All three Trilateral Offices concluded that their practices reflect consensus on two issues: "that a technical aspect is necessary for a computer-implemented business method to be eligible for patenting" and that "to merely automate a known human transaction process using well known automation techniques is not patentable" (Trilateral Technical Meeting 2000). With respect to the first point, a footnote observed that the USPTO permits the technical aspect to be implicit in the claim, whereas the EPO and the JPO require it to be explicit. With respect to the second, one wonders why the USPTO did not have it mind when it allowed Amazon.com's one-click on-line shopping patent. In any event, in light of this experiment, it is not surprising that one leading American source concludes that the JPO "appears to be following the lead of *State Street* in permitting patents on business methods" (Merges and Duffy 2002, 174).

Status of business methods under TRIPS. As noted in the introductory section, Article 27 of TRIPS requires WTO member countries to grant patents on "products or processes, in all fields of technology, provided that they are new, involve an inventive step and are capable of industrial application." Such patents must be granted, moreover, "without discrimination as to . . . the field of technology." The question has been raised whether these provisions amount to a command that WTO nations recognize business method patents or risk trade sanctions from those that do (Taketa 2002, 964–67).

This very argument was raised by the appellant in the EPO Boards of Appeal's 1998 *IBM* decision (see footnote 7), which dealt with a patent on a "computer program product." Interestingly, "[t]o a large extent the Board share[d] the appellant's opinion about the significance of TRIPS." The problem, however, was that the board, "for the time being," was "not convinced that TRIPS may be applied directly to the EPC," since it is an agreement among individual states. Despite its unwillingness to apply TRIPS directly, the board thought it "appropriate to take it into consideration." The board concluded "that it is the clear intention of TRIPS . . . not to exclude programs for computers as mentioned in and excluded under Article 52(2)(c) EPC." TRIPS, in other words, seems to require a *State Street* rule of patentable subject matter.

This is a rather startling statement: The EPC and TRIPS are in direct conflict. Presumably, if TRIPS did apply directly to the EPO, Article 52(2)(c) would be invalid and the EPO would have to adopt the U.S. approach. It is not clear how this could happen unless the EPC members were to decide to adopt TRIPS directly; such a fundamental revision of the EPC's text through indirect means seems unlikely at best. But in so-called monist countries, where treaties such as TRIPS immediately become part of national law, accession to TRIPS may already (albeit stealthily) have effected the adoption of the *State Street* regime.[9] The effect of TRIPS, whether direct or indirect, is a legal theme to be watched in the coming years.

Future Legal Trends

USPTO: Trends, practices, and initiatives. In March 2000, the USPTO announced a major plan "to improve the quality of the examination process in technologies related to electronic commerce and business methods" (USPTO 2000). In a white paper issued in conjunction with this announcement, the USPTO reviewed the history of business method patents as well as current trends and described several initiatives designed to add examiners, improve their competence, provide better access to relevant prior art, and insure quality control. Progress on these initiatives was the subject of a "Partnership Meeting" with USPTO "customers" in the summer of 2002 (USPTO 2002).

According to the white paper (USPTO 2000), the trend that was already in progress before the *State Street* decision has accelerated. For example, in Class 705 (data processing: financial, business practice, or cost/price determination), the USPTO received 330 applications in 1995, 584 in 1996, 927 in 1997, 1,340 in 1998, 2,821 in 1999, 7,800 in 2000, and an estimated 10,000 in fiscal year 2001. The number of allowed patents, which, of course, will lag behind applications, has also gone up steadily, from 203 in 1995 to 1,062 in 2000 although the most recent data for fiscal year 2001 suggest that allowances will drop into the 500 to 600 range. This drop is likely to be the result of more examiners giving each application greater scrutiny. The number of examiners in the work group that handles business method–related applications has almost doubled over the last two years, and those hired are said to have greater expertise in both business and computer applications. In addition, examiners are being furnished and encouraged to use wider resources for locating potentially disabling nonpatent prior art.

All allowed applications in Class 705 are now subjected to a second-level review "to ensure compliance with the mandatory search requirements, clarity and completeness of reasons for allowance, and to determine whether the scope of the claim should be considered." In addition, there is in-process review of randomly selected pending

9. Civil law countries commonly follow the monistic approach. Most common law countries are "dualist," meaning that national implementing legislation is required. The United States is a hybrid: the supremacy clause of Article VI of the Constitution provides that "all Treaties made, or which shall be made, under the Authority of the United States, shall be the supreme Law of the Land," but Congress regularly enacts implementing legislation (Taketa 2002, 960.)

cases.[10] Two measures of this heightened scrutiny are that the average time from filing to first office action (the initial notice of allowance and/or rejection of claims) in Class 705 is 23.5 months, versus 14.6 months for the entire USPTO, and that the average time to final disposition is 28.5 months in Class 705, versus 25.6 months for the entire USPTO (USPTO 2001).

It is difficult for an outsider to discern qualitative trends in the USPTO's response to business method patents. There is no reason to suspect that there has been any rearguard action against *State Street.* On the contrary, as noted earlier, the USPTO's *Guidelines* and *Manual* reflected skepticism about the vitality of the business method subject matter exception even before *State Street.* Its more recent documentation, both internal and external, is entirely faithful to the Federal Circuit's party line. All evidence points to the USPTO's much-publicized heightened scrutiny being focused instead on the categories of novelty and nonobviousness. Examiners are being instructed—and are being given better resources—to determine whether claimed applications really are new and nonobvious. The USPTO has clearly recognized that traditional searches in prior patents and professional literatures are inadequate to this task.

In the end, economics may have more to say than law about whether and when the business method patent flood finally crests.

This recognition does not mean that inventors have been—or are likely to be—deterred from filing highly aggressive patents. Nor has the USPTO ceased granting highly controversial patents. To cite just one example, on October 1, 2002, Ed Pool, owner of the one-room company DE Technologies, Inc., obtained U.S. Patent No. 6,460,020 on a "Universal Shopping Center for International Operation." The purpose of Pool's system is to provide "a pre-transactional calculation of all charges involved in any international transaction," including currency conversions, customs duties, freight, and insurance. The system is intended to do all the related paperwork electronically, in a language of the customer's choosing. Even before the patent issued, commentators speculated that it might be worth $2.4 billion in license fees from major Internet businesses; one predicted that "the patent will undoubtedly add to the uproar over business method patent policy" (Cronin 2000).

U.S. legislative prospects. The limits on enforcement of business method patents contained in the American Inventors Protection Act of 1999 have not satisfied some members of Congress. Subsequent sessions have seen the introduction of bills intended to impose even tighter restrictions. None has yet succeeded. Interestingly, the preferred approach has been not to attack *State Street* directly but to make the examination procedure more rigorous. Both H.R. 5634, the "Business Method Improvement Act of 2000" (106th Cong., 2d sess.), and H.R. 1322, the "Business Method Improvement Act of 2001" (107th Cong., 1st sess.), sought to give third parties and the public an enhanced opportunity to oppose business method patents as well as to raise the bar for novelty and nonobviousness. Both expired at the House subcommittee stage.

At the other extreme, some members of the patent bar have begun to argue for legislation to enhance enforceability. One recent article, for example, has pointed out that a U.S. business method patentee may be without a remedy under existing law "when the infringer has located part of the claimed process outside of the United States" and suggests ways in which Congress might "tweak the law" (Connor and Leak 2002, 1, 3). Even the authors concede, however, that there is no indication and little likelihood that Congress will act.

Future U.S. judicial issues. The Federal Circuit's opinion dissolving the preliminary injunction in the *Amazon.com* case (discussed earlier) may be the best predictor of future battles in the courts. The Federal Circuit focused not on the patentable subject matter issue but rather on novelty and nonobviousness. This is likely to be the pattern in cases to come: The subject matter issue has been laid to rest, and litigants will argue over how to apply Sections 102 and 103 in the business method field. One particular area to watch will be how the courts respond to the USPTO's increasing attention to nonpatent prior art. In most fields, novelty and nonobviousness litigation has focused overwhelmingly on previous patents. Relatedly, the USPTO has been making increasing use of "officially noted" subject matter to reject business method patents on novelty nonobviousness grounds. This subject matter, which does not involve statutory categories of prior art (other patents, printed publications, etc.)

at all, consists of "examples capable of instant and verifiable recognition, such as ATM machines" (USPTO 2002). Its relevance to business method applications is self-evident. As there is only limited case law dealing with officially noted subject matter (USPTO 2002), the judicial response will bear watching here as well.

Enforcement is another area in which de facto judicial limits on business method patents could emerge. Patent lawyers are sometimes heard to say that "what the Federal Circuit giveth on patentability, it taketh away on enforcement." They mean that, since its inception in 1983, the Federal Circuit has tended to be more expansive than the "secular" courts in allowing patents (see *State Street*, for example) but somewhat stingier in its willingness to find infringement.[11] The Supreme Court has contributed to this trend in recent years with two decisions that have narrowed the "doctrine of equivalents" (*Warner-Jenkinson Co. v. Hilton Davis Chemical Co.,* 520 U.S. 17 [1997]; *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 122 S. Ct. 1831 [2002]). This old, judicially created rule, whose purpose is to catch infringers who do not literally copy the patented invention, has shrunk almost to the point of disappearance. Perhaps this relative stinginess will creep into business method patent litigation.

A test of this hypothesis may be forthcoming. As noted in the previous section, some patent lawyers are suggesting that infringers may be able to take advantage of loopholes in the existing remedial section of the Patent Act (35 U.S.C. § 271) by conducting portions of their activities outside the United States (Connor and Leak 2002). Some lawyers also suggest that the Federal Circuit might close those loopholes by aggressive construction of the act. We will know more if and when these loopholes begin to be litigated.

EU initiatives. Within the European Union, two ongoing developments are noteworthy. First, the creation of a true EU patent continues to be a major priority. Motivated by concerns about the transaction costs of the EPO's national patent bundle scheme, as well as the inconsistencies that can result from national enforcement, the European Commission proposed the creation of a Community Patent in July 2000. In a May 2002 speech (EU 2002), EU Internal Market Commissioner Frits Bolkestein indicated that although "the Council has made progress on making such a Community Patent a reality," there were still substantial roadblocks in the form of "the interests of a small number of specialists, judges, and lawyers that currently work in national patent courts." However, press reports in March 2003 indicated that the EU ministers had reached a compromise on these remaining issues, paving the way to finalization of the EU patent.

At the same time, the EU Commission has presented a proposal for a Directive on the Patentability of Computer-Implemented Inventions (Comm. of the EC 2002). If adopted by the European Parliament and Council, commission directives require the harmonization of member state laws in accordance with their contents. The overall thrust of this directive is to solve the inconsistent enforcement problem by requiring the adoption of the EPC standards as the national law of the EU member states.

Under the current official text of the proposed directive, computer-implemented inventions are "considered to belong to a field of technology" (Art. 3). Like other inventions, they must meet the traditional European standards of industrial application, novelty, and inventive step. In order to meet the inventive step requirement, a computer-implemented invention "must make a technical contribution" (Art. 4[2]). This requirement means that the nonobvious contribution to the art must be in a technical area, whether it lies in the underlying problem, the solution, or the effects of the solution. Significantly, "if there is no technical contribution, e.g., if the contribution to the state of the art lies wholly in non-technical aspects, as would be the case if the contribution to the state of the art comprised purely a method of doing business, there will be no patentable subject matter" (p. 14 [Explanation of the Directive: Article 4]). Nonetheless, if the technical contribution requirement is met, the claim "may comprise both technical and non-technical features" (Art. 4[3]), meaning that the scope of the patent will not be limited to the technical contribution. The EU approach is consistent with the EPC, EPO *Guidelines*, and cases such as *Sohei,* which prohibit the patenting of business methods "as such" but find an ultimate business

10. At the 2002 Partnership Meeting, a participant asked why there was no second-level review of disallowed applications: Is the USPTO "telling examiners that they can do low quality examination for cases they do not want to allow?" (USPTO 2002). The USPTO's response was to cite the random in-process review initiative.
11. See, for example, *Wang Laboratories, Inc. v. America Online, Inc.* (197 F.3d 1377 [Fed. Cir. 1999]), in which the court narrowly construed the claims in a software-based invention so as to affirm a lower court finding of no infringement. The Federal Circuit also declined to apply the doctrine of equivalents, discussed in the text above.

method objective to be irrelevant so long as the claimed invention is suitably technical in character.

UKPO review. Like its counterparts in the United States and Japan, the UKPO has recently reviewed its policies concerning business method patents (UKPO 2001). Its request for consultation with its various constituencies produced 285 formal submissions by both individuals and organizations and 11,000 Web site hits. Not surprisingly, "[t]here was no consensus among respondents on how far software ought to be patentable" (¶11). With respect to software, the UKPO's position was "to reaffirm the principle that patents are for technological innovations. Software should not be patentable where there is no technological innovation, and technological innovations should not cease to be patentable merely because the innovation lies in software" (¶19). This conclusion seems consistent with the EPO position and, arguably, somewhat more restrictive than the *State Street* rule. With respect to business methods, the UKPO concluded that those advocating patentability "have not provided the necessary evidence that it would be likely to increase innovation. Unless and until that evidence is available, ways of doing business should remain unpatentable" (¶24). The latter position is consistent with that taken by the British courts. It is thus less favorable to patents than the position taken by the EPO and, of course, the USPTO and the U.S. Federal Circuit.

Conclusion

There is at present a substantial rhetorical gap between the United States on the one hand and Europe and Japan on the other concerning the patentability of business methods. Under the *State Street* decision, business methods clearly constitute patentable subject matter. Europe and Japan, with their "technical character" requirement, see themselves (as reflected in their official literatures) as imposing significant barriers to patentability both at the subject matter and inventive step stages of the examination. They view the United States as having come down strongly and perhaps irrevocably in favor of ready patentability. The EU, for example, has characterized the United States as a "test case," conducting a potentially dangerous experiment with its "negligible" restrictions on business method patents (Comm. of the EC 2002, 5).

The theme that emerges from this paper, however, is that the differences that seem so striking at the theoretical level might not be so profound in practice. While the Federal Circuit has forced the USPTO to renounce subject matter objections to business method patents (something it was probably on the way to doing anyway), the USPTO has taken significant steps to scrutinize novelty and nonobviousness more rigorously. In its *Amazon.com* decision, the Federal Circuit showed an inclination to do the same. Thus, while the American patent system may be perceived abroad as having given a blank check to business method applicants, the reality may prove to be considerably more restrictive.

Europe and Japan, by contrast, may in practice be somewhat more liberal than their policy pronouncements would indicate. In the EPO, cases such as *Sohei* and the two *IBM* decisions suggest that the technical character barrier might be a matter more of form than substance, at least at the subject matter stage. And in Japan, the trilateral experiment revealed no differences with the United States in examination outcomes. Although the inventive step distinction remains material, the eventual outcome may nonetheless be convergence, with the United States turning out to be permissive in theory but perhaps demanding in practice, while Europe and Japan display precisely the opposite tendency. In the end, economics may have more to say than law about whether and when the business method patent flood finally crests.

REFERENCES

Blumenthal, David A., and Bruce D. Riter. 1980. Statutory or non-statutory? An analysis of the patentability of computer related inventions. *Journal of the Patent Office Society* 62:454–520.

Chisum, Donald S. 2002. *Chisum on patents.* Newark, N.J.: LexisNexis.

Commission of the European Communities (Comm. of the EC). 2000. The patentability of computer-implemented inventions. Consultation Paper by the Services of the Directorate General for the Internal Market, Brussels, October 19.

———. 2002. Proposal for a directive of the European Parliament and of the Council on the Patentability of Computer-Implemented Inventions. COM(2002) 92 final, 2002/0047 (COD), Brussels, February 20.

Connor, Michael S., and Frank W. Leak. 2002. Challenges of business method patent enforcement—extraterritoriality. *The Computer and Internet Lawyer* 19, no. 8:1–4.

Cronin, Kevin. 2000. Possible patent windfall. *Boston College Intellectual Property & Technology Forum* 2000:092701. <www.bc.edu/bc_org/avp/law/st_org/iptf/headlines/content/2000092702.html> (July 1, 2002).

European Patent Office (EPO). 2000. Patentability of methods of doing business. Press release. <www.european-patent-office.org/news/pressrel/2000_08_18_e.htm> (February 13, 2003).

European Union (EU). 2001. The economic impact of patentability of computer programs. Internal Market Study. <www.europa.eu.int/comm/internal_market/en/indprop/comp/studyintro.htm> (December 27, 2002).

———. 2002. Results of Internal Market, Consumer Affairs, and Tourism Council, Brussels, May 21, 2002. Internal Market Study. <www.europa.eu.int/comm/internal_market/en/indprop/patent/imc/05-02res_en.htm> (October 16, 2002).

Japanese Patent Office (JPO). 2000. Examination Information: Policies concerning “Business Method Patents.” <www.jpo.go.jp/tetuzuki_e/index.htm> (December 15, 2003).

Kesan, Jay P. 2002. Carrots and sticks to create a better patent system. *Berkeley Technology Law Journal* 17:764–97.

Kuester, Jeffrey R., and Lawrence E. Thompson. 2001. Risks associated with restricting business method and e-commerce patents. *Georgia State University Law Review* 17:657–68.

Merges, Robert P., and John F. Duffy. 2002. *Patent law and policy: Cases and materials.* Newark, N.J.: LexisNexis.

Meurer, Michael J. Forthcoming. Business method patents and patent floods. *Washington University Journal of Law and Policy.*

Oxford Intellectual Property Research Centre (IPRC). 2000. The first mover monopoly: A study on patenting business methods in Europe. Olswang and Oxford Intellectual Property Research Centre, Oxford University. Oxford IPRC Working Paper 05/00. <www.oiprc.ox.ac.uk/EJWP0500.html> (February 21, 2003).

Taketa, Jason. 2002. Notes: The future of business method software patents in the international intellectual property system. *Southern California Law Review* 75 (May): 943–82.

Trilateral Technical Meeting. 2000. Comparative study carried out under Trilateral Project B3b. Trilateral Technical Meeting, Japanese Patent Office, Tokyo, June 14–16. <www.uspto.gov/web/tws/front_page.pdf> (December 15, 2003).

U.K. Patent Office (UKPO). 2001. Should patents be granted for computer software or ways of doing business? The government’s conclusions. <www.patent.gov. uk/about/consultations/conclusions.htm> (July 1, 2002).

U.S. Patent and Trademark Office (USPTO). 2000. Automated financial or management data processing methods (business methods). USPTO White Paper. <www.uspto.gov/web/menu/busmethp/index.html> (November 11, 2002).

———. 2001. Business methods still experiencing substantial growth—report of fiscal year 2001 statistics. <www.uspto.gov/web/menu/pbmethod/fy2001strport.html> (November 11, 2002).

———. 2002. Business Methods Summer Partnership Meeting follow up questions. <www.uspto.gov/web/menu/pbmethod/summer2002qanda.html> (November 11, 2002).

Take Your Model Bowling: Forecasting with General Equilibrium Models

MARCO DEL NEGRO AND FRANK SCHORFHEIDE

Del Negro is a research economist and assistant policy adviser in the Atlanta Fed's research department. Schorfheide is an assistant professor of economics at the University of Pennsylvania. *They thank Eric Leeper, Ellis Tallman, and Rochelle Edge for helpful suggestions.*

In the past twenty years dynamic stochastic general equilibrium (DSGE) models have taken center stage in academic macroeconomic research. The stated goal of DSGE models in the tradition of Kydland and Prescott (1982) is to explain business cycle features of the data and to be usable for quantitative—as opposed to only qualitative—policy analysis. Yet until recently the data that these models have been measured against are not the GDP or inflation figures that appear in newspapers but so-called filtered data. One commonly used filter, the Hodrick-Prescott (1997) filter, decomposes the data into a cyclical component and a growth component and removes the latter. By doing so the filter removes from the data variations that are due to frequencies other than business cycle frequencies. One rationale behind the filtering is that the model is designed to explain business cycles as opposed to very short-run (say, seasonal) or long-run movements (say, due to demographics) in the data. Hence, it seems logical to assess the model's fit in terms of that part of the data that it can explain.

Whatever the motivation behind using filtered data, filtering has two important consequences. First, it implies that the task of forecasting macroeconomic time series stays outside the realm of DSGE models and is left entirely to econometric models or judgmental forecasters. Practitioners are interested in forecasts of actual, as opposed to filtered, data, so they rely on models, or individuals, that deliver such forecasts. The second consequence of using filtered data is that, to this day, policymakers rarely use general equilibrium models, at least in quantitative analysis. Like practitioners, policymakers base their decisions on forecasts of macroeconomic time series. Policymakers want to know, for instance, the expected path of inflation, unemployment, or real output growth in the next few quarters and by how much a 25 basis point cut in the federal funds rate would change such a path. Since very little is known about the forecasting performance of general equilibrium models, policymakers rarely rely on them for quantitative policy assessment.

Many of the models currently used in forecasting and policy analysis belong to one of two categories. The first includes models in the Cowles Commission tradition.[1] These are large-scale simultaneous equation models that were prominent in macroeconomics before the rational expectations revolution, from the late 1950s to the early 1970s (see Diebold 1998 for a brief history of macroeconomic forecasting). These models have been updated to incorporate rational expectations and are still heavily used for forecasting and policy-making by central banks around the world as well as by commercial forecasters.[2] FRB/US—the workhorse model of policy analysis at the Federal Reserve Board of Governors—is one of them.[3] The second category of models includes vector autoregressions (VARs), which were introduced by Sims (1980) in the early 1980s and popularized in the forecasting literature by Litterman (1986). The BVAR model (a VAR with Bayesian priors) used for forecasting at the Federal Reserve Bank of Atlanta belongs to this second category.[4]

Like VARs and Cowles Commission models, DSGE models also aspire to describe the data. Perhaps the main difference between DSGE models on the one side and VARs and Cowles Commission models on the other side is that DSGE models are explicitly derived from first principles. That is, DSGE models describe the general equilibrium allocations and prices of a model economy in which agents (households, firms, financial intermediaries, etc.) dynamically maximize their objectives (utility, profits, and so on) subject to their budget and resource constraints. The DSGE model parameters describe the preferences of agents (tastes), the production function (technology), and other features of the economy. These parameters are called "deep" parameters—parameters that do not vary with policy.

The stated goal of DSGE models in the tradition of Kydland and Prescott (1982) is to explain business cycle features of the data and to be usable for quantitative policy analysis.

To be sure, economic theory also informs Cowles Commission–style models and VARs. For instance, most equations in Cowles Commission–style models, such as consumption equations, investment equations, and so on, are inspired by economic analysis, if not explicitly derived from it. However, in some cases the parameters of these models characterize behavior instead of tastes and technologies. Yet the agents' behavior is not policy invariant, and therefore not all parameters in such models are deep.[5] The modelers typically adopt a block-by-block approach (in which the blocks are the household sector, the business sector, etc.; see Brayton, Levin, et al. 1997; Brayton, Mauskopf, et al. 1997) to describe the various agents and sectors in the economy and often ignore important links among blocks. In particular, when forming expectations, agents in these models often ignore equilibrium restrictions that must hold in all future states of the world. VARs were introduced by Sims (1980) with the intent to overcome the deficiencies of the Cowles Commission approach and to obtain specifications that are consistent with a dynamic general equilibrium. In fact, a linearized DSGE model can be closely approximated by a VAR with a sufficiently large number of lags. The VAR parameters can, in principle, be constrained to be functions of deep parameters for some DSGE model. Typically, however, VARs are not estimated under such constraints, and therefore the VAR parameter estimates cannot be interpreted in terms of deep parameters.

This article reviews some recent attempts to use general equilibrium models for forecasting and policy analysis. In particular, the article focuses on one specific approach, pioneered by Ingram and Whiteman (1994) and further developed by Del Negro and Schorfheide (forthcoming), that relies on the use of general equilibrium models as priors for Bayesian VARs. To motivate this approach, which we will call DSGE-VAR, we first need to address two questions. First, why should one bother to forecast with general equilibrium models? Second, why should one use general equilibrium models as priors instead of forecasting directly with them? The next two sections address these questions.

Why Forecast with DSGE Models?

There are two good reasons to use DSGE models in forecasting (also see Diebold 1998 for a discussion of forecasting with DSGE models). The first reason has to do with improving the forecasting precision. It is well known that loosely parameterized models, such as VARs, are imprecisely estimated unless a very long time series of data is available, which is rarely the case in macroeconomics. Imprecise estimates in turn result in potentially large forecast errors, especially for long forecast horizons. A solution to this problem of too many parameters is to use Bayesian priors. In Bayesian econometrics a *prior* on a set of parameters is a distribution that summarizes beliefs or knowledge about these parameters prior (whence the name) to observing the data. Priors reduce the sample variability in the parameter estimates by "shrinking" them toward a specific point in the parameter space. For this reason, since the seminal work of Litterman (1986) and Doan, Litterman, and Sims (1984), BVARs have earned a reputation for forecasting accuracy (see Robertson and Tallman 1999 for a review of the comparative forecasting accuracy of BVARs). In many BVARs the priors arise from statistics, namely, from the observation that random walk processes describe quite well the behavior of a number of macroeconomic time series.[6] This observation is the rationale, for instance, for the well-known Minnesota prior. The Minnesota prior shrinks the VAR parameters toward a unit root. Ingram and Whiteman (1994) proposed to use a prior that comes from a general equilibrium model, namely, a standard real business cycle (RBC) model. Ingram and Whiteman show that the performance of their VAR with an RBC prior in terms of forecasting real variables (real output, consumption, and investment growth) is comparable to that of a VAR with a Minnesota prior.

The second reason for forecasting with DSGE models has to do with evaluating the impact of changes in policy. The well-known Lucas (1976) critique implies that only models in which the parameters are deep—that is, models in which the parameters do not vary with policy—are suited to evaluate the impact of policy changes. To understand why this is the case, let us consider a model in which the parameters are not deep; this may be a VAR or a Cowles Commission model. The forecaster who uses such a model to predict the effect of a given policy change faces the following dilemma. On the one hand, she can estimate the parameters of the model only on the basis of past data and experiences. On the other hand, unless the policy change has occurred before, she can gain little guidance from past experience about how the policy change affects the decision rules of agents and hence how it affects the parameters of the model.

For example, suppose that the goal is to predict the effects of the 2003 change in the tax code. A forecaster might use the data available prior to the policy change to estimate a consumption equation that describes the behavior of consumers as a function of a number of variables, including wealth and disposable income. Knowing the amount by which wealth and disposable income will be increased by the tax breaks, the forecaster may use the estimated relationship to forecast consumption in the next few quarters. However, the Lucas argument is that the change in policy may induce agents to change their behavior, which in turn may change the relationship between wealth, disposable income, and consumption. For instance, the tax break may affect the agents' propensity to consume. Hence, the forecast for consumption may well turn out to be wrong.

Now suppose that the model being used to forecast the impact of the policy change is a DSGE model. In a DSGE model the parameters are truly deep, that is, invariant with policy, or at least they are assumed to be so. For instance, there is no reason to think that a change in the tax policy would affect either the extent to which people enjoy leisure (tastes) or the current speed of computers (technology).[7] Therefore the forecaster can estimate these parameters using existing data and does not have to worry that they may change with policy. Once the parameters are available, the forecaster can solve the model and work out the impact of the tax change on consumption. For instance, the forecaster using DSGE models can correctly compute agents' propensity to consume under the new policy. If the specification of the DSGE model is appropriate, the effect of the new policy can be correctly evaluated even though it has not occurred in the past.

Of course, the distinction just drawn between models with and without deep parameters is Manichaean. First, not all parameters in DSGE models are necessarily deep, so some DSGE models may be subject to the same criticism as the other models. Second, not all policy changes result in dramatic changes in agents' behavior. In such cases, models other than DSGE models may well be able to provide reliable forecasts.[8] With these important caveats established, one of the main implications of the Lucas critique is that DSGE models have in principle an important advantage over other models in forecasting the effects of policy changes.

Why Use Priors?

The advantages of DSGE models discussed in the previous section often come at a cost in terms

1. The Cowles Commission (now the Cowles Foundation) was founded by Alfred Cowles in 1932 to promote quantitative research in economics. The Cowles Commission and its fellows played a pivotal role in promoting and developing large-scale econometric models. Hence, its name became associated with the approach (see Fair 1992).
2. Sims (2002) provides a criticism of the way Cowles Commission–style models are currently being estimated and used for policy analysis.
3. See Brayton, Mauskopf, et al. (1997) and Brayton, Levin, et al. (1997) for a description of FRB/US and Reifschneider, Stockton, and Wilcox (1997) for a description of forecasting and policy evaluation at the Federal Reserve Board of Governors.
4. See Zha (1998) for a discussion of the use of VARs in policy analysis.
5. Production functions and utility functions underpin the equations for many sectors of models like FRB/US, however. As a result, one may be able to back out some of the deep parameters from the coefficient estimates.
6. A random walk is a process in which today's best guess about tomorrow's value of a variable is today's value, possibly augmented by a constant.
7. This assumption is not true for all tax policy changes: Some changes may affect spending on research and development and therefore future technology. However, as a first approximation this effect can be ignored for many policy changes.
8. See Sims (1982), Leeper, Sims, and Zha (1996), and Sargent (1984) for a discussion of the relevance of the Lucas critique for VARs.

of the model's fit. A number of papers that study the fit of DSGE models (for example, Altug 1989; Leeper and Sims 1994; Ireland 1997; Schorfheide 2000) find that this fit is far from perfect. As discussed above, economic theory imposes a number of restrictions on the stochastic process followed by the data—the cross-equation restrictions that are the hallmark of rational expectations econometrics. These restrictions imply that DSGE models are scarcely parameterized compared with VARs or Cowles Commission–style models. Hence, DSGE models may match the data in many important dimensions but, being overly simplified model economies, may also fail in several other dimensions, resulting in large forecast errors for some of the variables of interest.

One of the main implications of the Lucas critique is that DSGE models have in principle an important advantage over other models in forecasting the effects of policy changes.

If one wants to forecast with general equilibrium models, using them indirectly as priors may be preferable to the alternative approach of forecasting directly with them. Using general equilibrium models as priors means that the restrictions stemming from economic theory are imposed loosely instead of rigidly. This method implies that the final (posterior) stochastic process used to forecast will respect the restrictions in those dimensions where these restrictions are not rejected by the data but may otherwise depart from them. Of course, if the restrictions are too loosely imposed there is virtually no difference between forecasting with general equilibrium priors and with an unrestricted VAR (a VAR without priors). Therefore, a key input in the process is the "degree of tightness"—which will be denoted hereafter as λ and which ranges from 0 (no prior) to ∞ (rigid restrictions). The remainder of the article will describe how to choose λ optimally.

How Does DSGE-VAR Work?

The discussion in this section offers an intuitive exposition of the procedure in Del Negro and Schorfheide (forthcoming) for using DSGE models as priors in VARs. Assuming that T observations are available for the variables to be forecast (for instance, real output growth, inflation, and the short-term interest rate), the procedure amounts to generating λT observations for real output growth, inflation, and the federal funds rate from the DSGE model; combining these dummy observations with the actual data; and running a VAR on the augmented data set.[9]

The DSGE-VAR procedure assumes that the dynamics for the data to be forecast are reasonably well described by an unrestricted VAR. To the extent that these dynamics are linear and that the VAR has a sufficient number of lags, this is not a very heroic assumption: VARs are parameterized loosely enough to accommodate nearly any linear stochastic process. As discussed above, the problem is precisely that VARs have too many parameters, so the estimates may be imprecise in short samples. Our approach starts from the premise that a DSGE model may provide useful restrictions for the VAR parameters—useful in the sense that the restrictions can improve the model's forecasting performance. We do not want to impose these restrictions dogmatically for the reasons described in the previous section. Rather, we treat the DSGE model as prior information in the estimation. As is well known since the work of Theil and Goldberger (1961), one way to incorporate prior information into the estimation is to augment the sample with dummy observations that reflect the prior (see also Sims and Zha 1998). This is precisely the route we follow: Our dummy observations are simply data generated by the DSGE model.

The next step in the procedure consists of estimating the VAR parameters using both the actual and the dummy observations. To make this step clear, we specify some notation. Y is the $T \times n$ matrix of actual data, where T is the sample size and n is the number of variables. X is the matrix of VAR regressors, which includes the constant as well as the lags of the variables. The VAR, which we assume to be the data-generating process, is given by

$$(1)\quad Y = X\Phi + U,$$

where U is the $T \times n$ matrix of VAR innovations, which are normally distributed with mean 0 and variance Σ_u. The standard OLS estimator for Φ is given by the well-known formula

$$(2)\quad \Phi_{OLS} = (X'X)^{-1}X'Y.$$

Now λT observations are generated for the variables of interest from the DSGE model. As mentioned above, λ is the weight of the prior. So if $T = 100$ and $\lambda = 0.5$, this step generates fifty observations from the DSGE model. Using these dummy observations, the matrices Y^* and X^* are constructed. Finally, the OLS estimates are run again on the augmented dataset that includes both actual and dummy observations, yielding the estimator

$$(3)\quad \Phi_{DSGE\text{-}VAR} = (X'X + X^{*\prime}X^*)^{-1}(X'Y + X^{*\prime}Y^*).$$

This estimator will be used in forecasting.

Now that the formula is determined, a few comments are in order. First, we want to elaborate on the role of λ, the weight of the prior. Notice that the previous formula can be equivalently expressed as

$$(4)\quad \Phi^{\lambda}_{DSGE\text{-}VAR} = \left(\frac{1}{1+\lambda}\frac{X'X}{T} + (1-\frac{1}{1+\lambda})\frac{X^{*\prime}X^*}{\lambda T}\right)^{-1} \left(\frac{1}{1+\lambda}\frac{X'Y}{T} + (1-\frac{1}{1+\lambda})\frac{X^{*\prime}Y^*}{\lambda T}\right)$$

The terms $(X'X)/T$ and $(X'Y)/T$ are the second moments (that is, say, the covariance between real output growth today and interest rates in the previous period) computed from the data. The terms $(X^{*\prime}X^*)/\lambda T$ and $(X^{*\prime}Y^*)/\lambda T$ are the second moments implied by the DSGE model. Our proposed estimator is computed by weighting the second moments from the data with the second moments implied by the DSGE model, with weights that are respectively $1/(1 + \lambda)$ and $1 - 1/(1 + \lambda)$. If $\lambda = 0$, the dummy observations disappear from the formula: Since for $\lambda = 0$ we are using only the second moments from the data, the estimator in this case coincides with the OLS estimator. If $\lambda = \infty$, the weight on the dummy observations becomes 1. Thus, for $\lambda = \infty$ the restrictions coming from the DSGE model are rigidly imposed.

Next, we introduce an important refinement into the procedure. Whenever λT is not too large (say, $\lambda = 1/10$, and $T = 100$), we generate a small number of dummy observations (in the above example, $\lambda T = 10$). Because of sample variability in the Monte Carlo procedure that generates the dummy observation from the DSGE model, whenever λT is small the *sample* second moments, the terms $(X^{*\prime}X^*)/\lambda T$ and $(X^{*\prime}Y^*)/\lambda T$, may provide a poor estimate of the *population* second moments that the DSGE model implies. One way around the problem is to compute the terms $(X^{*\prime}X^*)/\lambda T$ and $(X^{*\prime}Y^*)/\lambda T$ a large number of times and then average across realizations. This way of proceeding has the disadvantage of being computationally expensive because one would have to draw over and over from the DSGE model. We follow an alternative approach and exploit the fact that whenever the DSGE model is linear (or is well approximated by a linear solution) the population second moments, which we call Γ^*_{xx} and Γ^*_{xy}, can be computed analytically. Hence, in the formula for the estimator $\Phi_{DSGE\text{-}VAR}$, we use the population moments Γ^*_{xx} and Γ^*_{xy} in place of $(X^{*\prime}X^*)/\lambda T$ and $(X^{*\prime}Y^*)/\lambda T$.

Up to this point we have not mentioned the values taken by the deep parameters (preferences, etc.) of the DSGE model. We denote with θ the vector of deep parameters. Clearly, the population moments Γ^*_{xx} and Γ^*_{xy}, and hence our estimator $\Phi_{DSGE\text{-}VAR}$, will depend on the choice of θ. To make this dependence explicit, we rewrite the estimator as

$$(5)\quad \Phi^{\lambda}(\theta)_{DSGE\text{-}VAR} = \left(\frac{1}{1+\lambda}\frac{X'X}{T} + (1-\frac{1}{1+\lambda})\Gamma^*_{xx}(\theta)\right)^{-1} \left(\frac{1}{1+\lambda}\frac{X'Y}{T} + (1-\frac{1}{1+\lambda})\Gamma^*_{xy}(\theta)\right).$$

In the macro literature that follows Kydland and Prescott (1982), a popular approach for choosing θ is calibration (see Kydland and Prescott 1996). Calibration amounts to selecting the values of θ on the basis of information other than that contained in the data we want the model to explain (or forecast). This information may come from microeconomic studies as well as from long-run empirical relationships, such as the labor share of national income or the consumption-output ratio.

We choose to depart from calibration and estimate θ; that is, we let the value of θ be determined by the data we want to fit. We do so on the grounds that if the calibration exercise is poorly performed, or if there is little outside information to pin down some of the elements of θ, the forecasting performance of DSGE-VAR may be severely affected. Still, in order to take advantage of useful outside information (micro studies and so on), we incorporate prior information into the estimation of θ.

How do we learn about θ from the data in our procedure? From equation (1), the data depend on the VAR parameters, not on θ. The answer is that we learn about θ indirectly, via the estimator $\Phi_{DSGE\text{-}VAR}$. As emphasized in equation (5), as long as λ is greater than zero, the choice of θ affects $\Phi_{DSGE\text{-}VAR}$. From the data we learn which $\Phi_{DSGE\text{-}VAR}$ has the best fit. But since for each choice of $\Phi_{DSGE\text{-}VAR}$ there corresponds a choice for θ, we can go back and learn from the data about θ. Note that whenever $\lambda = \infty$,

9. See Del Negro and Schorfheide (forthcoming) for an econometrically detailed description of the approach as well as an appendix on how the procedure works in practice.

that is, whenever the restrictions coming from the DSGE model are imposed rigidly, our estimator for θ coincides with Smith's (1993) SQML (simulated quasi-maximum likelihood) estimator.

How Much Should the DSGE Prior Matter?

The discussion in the previous section emphasized that the choice of λ is crucial in the estimation. Our procedure does not require the forecaster to have strong a priori views on the choice of λ—that is, to pick λ ex ante. Rather, as the forecaster learns about θ from the data, she can also learn about λ. This section shows how λ can be estimated endogenously.

DSGE-VAR addresses regime shifts, trying to strike a balance between the forecasting accuracy of BVARs and the compliance to the Lucas critique of DSGE models.

The intuition about how to choose λ is the same as the one given in the previous section on the estimation of θ. Again, the data do not depend on λ but only on the VAR parameters. However, as formula (5) shows, the estimator $\Phi_{DSGE\text{-}VAR}$ crucially depends on the choice of λ. To make this explicit, let us write $\Phi^{\lambda}_{DSGE\text{-}VAR}$ (for simplicity, in this section we abstract from the choice of θ, which we can think of as fixed). If λ tends to infinity, the resulting estimator $\Phi^{\infty}_{DSGE\text{-}VAR}$ will conform to the restrictions imposed by the DSGE model. Otherwise, it will not. To the extent that the restrictions coming from the DSGE model lead to an estimator that fits the data well, the procedure points toward choosing a high value for λ.

In the above discussion, the definition of "fit" must be clarified. Fit does not simply correspond to large values of the likelihood function or small values of the in-sample sum-of-squared residuals. It is clear that the unrestricted estimator ($\lambda = 0$) always beats the restricted estimator ($\lambda > 0$) in terms of in-sample fit: A constrained optimum cannot fare any better than the unconstrained optimum. What we have in mind is the fit of the model, taking into account the model complexity. Consider the problem of choosing the lag length for a regular VAR. A popular criterion for lag-length selection is the Schwarz criterion. It penalizes the maximized likelihood function by a measure of model complexity, which is a function of the number of parameters to be estimated. The penalty term avoids the problem of the data being overfitted. The choice of λ works similarly, except that complexity cannot be determined by a simple parameter count (the number of VAR parameters is the same for all values of λ). We measure complexity as the degree of uncertainty associated with the parameter estimates. For instance, for $\lambda = 0$ the resulting estimator $\Phi^{0}_{DSGE\text{-}VAR}$ coincides with Φ_{OLS}, which is fairly imprecise. In this case we are using a large penalty. The higher λ, the more the estimator $\Phi^{\lambda}_{DSGE\text{-}VAR}$ is pulled toward the restrictions imposed by the DSGE model and the lower its variance. Hence, for large λ, the penalty that is used to adjust the measure of in-sample fit is small. Overall, if the DSGE model restrictions are very much at odds with the data, one would prefer the uncertainty and choose a low λ. If, however, the model is good, in the sense that the restrictions it imposes are not grossly at odds with the data, then one may welcome the reduction in uncertainty and choose a high value for λ.

In Bayesian terminology, our measure of fit coincides with the marginal likelihood, which is the integral of the likelihood function over all possible parameter values, weighted by the prior density. The marginal likelihood can be approximated by a penalized likelihood function as described above. We use the (exact) marginal data density to find the optimal value of λ (see Del Negro and Schorfheide, forthcoming, section 3.3.1).

The DSGE Model Used to Generate the Artificial Data

The methodology behind DSGE-VAR is general, so it does not depend on the specific DSGE model that is chosen. Of course, the better the DSGE model, in the sense that it captures the important features of the economy, the higher the weight λ it should receive in the composition of the augmented sample. We apply our procedure to a fairly standard and simple neo-Keynesian DSGE model. This section very briefly describes the model (see Del Negro and Schorfheide, forthcoming, section 2, for further details).

When written in log-linearized form (that is, all the variables are expressed in percentage deviations from their stochastic steady state), the model boils down to the following three equations:

1. an IS curve relating real output (x_t) to the level of the real interest rate, computed as the nominal rate minus expected inflation ($R_t - E_t\pi_{t+1}$), as well as to technology shocks (z_t), government

spending shocks (g_t), and expectations of future real activity ($E_t x_{t+1}$):

$$(6) \quad x_t = E_t x_{t+1} - \tau^{-1}(R_t - E_t \pi_{t+1}) + (1 - \rho_g)\, g_t + \rho_z \tau^{-1} z_t,$$

where τ, ρ_g, and ρ_z measure the agents' relative risk aversion and the degree of persistence of government and technology shocks, respectively;

2. a Phillips curve relating current inflation (π_t) to expectations of future inflation ($E_t \pi_{t+1}$), output, and government spending:

$$(7) \quad \pi_t = \beta E_t \pi_{t+1} + \kappa(x_t - g_t),$$

where κ measures the slope of the Phillips curve and is a function of deep parameters of the model; and

3. a Taylor rule, by which the monetary authority reacts to deviations of inflation from target and of output from potential output when setting the interest rate, R_t:

$$(8) \quad R_t = \rho_R R_{t-1} + (1 - \rho_R)(\psi_1 \pi_t + \psi_2 x_t) + \varepsilon_{Rt},$$

where ρ_R is the degree of persistence of monetary shocks and where the coefficients ψ_1 and ψ_2 represent the sensitivity of interest rates to output and inflation.

Forecasting with DSGE-VAR

The introduction to this article stressed the importance of forecasting, both for practitioners and policymakers. This section investigates the forecasting performance of DSGE-VAR in terms of three of the variables that most interest monetary policymakers: real output growth, inflation, and the federal funds rate. The reader must bear in mind that the results presented in this section are particular to the specific DSGE model described in the previous section. More elaborate models may generate different—and possibly better—results. Nonetheless, it is important to assess how DSGE-VAR fares when applied to a very simple model, if only for comparison.

In this section we specifically address two questions. First, is the DSGE prior useful in terms of forecasting? In other words, does the presence of the DSGE prior increase the forecasting performance relative to that of an unrestricted VAR? This question amounts to asking whether the restrictions that the DSGE model imposes on the VAR do good or harm when forecasting. Table 1 addresses this question and shows that by and large the DSGE prior is useful in terms of forecasting over a VAR with no priors. The table shows the percentage improvement in forecast accuracy relative to an unrestricted VAR for horizons from one to sixteen quarters ahead.[10] The forecast accuracy is measured as the root mean squared error of the forecast using a rolling sample from 1975Q3 to 2003Q3, a period that includes a number of recessions. At each point in the rolling sample, we estimate the model using eighty observations (say, at the first date in the rolling sample, we use data from 1955Q4 to 1975Q3; at the second date we use data from 1956Q1 to 1975Q4, etc.).[11] The importance of the DSGE prior, λ, is chosen optimally as described earlier. Of course, in principle the optimal λ depends on the sample—that is, it might change as we move from the beginning to the end of the rolling sample. In practice, as expected, the optimal λ was fairly constant over the rolling sample, around 0.5. A value for λ of 0.5 means that we used half as many artificial observations from the DSGE model as the number of actual observations.

The numbers in Table 1 are positive whenever the accuracy of DSGE-VAR is greater than that of the unrestricted VAR. One can readily see that the

TABLE 1

Percentage Gain in Root Mean Squared Error: DSGE-VAR versus VAR

Horizon (quarters)	Real GDP growth	Inflation	Federal funds rate
1	17.4	8.4	7.3
2	17.0	7.2	5.0
4	15.1	8.8	5.0
6	14.1	10.5	6.6
8	12.4	11.5	8.4
10	14.4	12.3	8.2
12	15.1	12.6	6.4
14	16.2	13.0	6.1
16	19.1	13.2	5.8

Note: The rolling sample is 1975Q3 to 1997Q3 (ninety periods). At each date in the sample, eighty observations are used to estimate the VAR. The forecasts are computed based on the optimal value of λ chosen ex ante.

10. For real output growth and inflation, the quantities being forecast are cumulative. In other words, for a sixteen-quarter horizon we are trying to forecast the average real output growth in the next four years as opposed to the rate of growth of the economy exactly sixteen quarters from now. Results obtained using the noncumulative forecasts deliver the same conclusion, however.

11. Other details of the exercise, such as the prior used for θ, are described in Del Negro and Schorfheide (forthcoming).

TABLE 2

Percentage Gain (Loss) in Root Mean Squared Error: DSGE-VAR versus BVAR with Minnesota Prior

Horizon (quarters)	Real GDP growth	Inflation	Federal funds rate
1	1.1	1.7	–7.6
2	7.0	1.3	–4.9
4	5.8	4.8	–1.9
6	3.5	7.2	–0.7
8	4.2	7.8	–0.2
10	8.0	8.4	–0.6
12	12.5	9.0	0.7
14	17.2	9.6	1.1
16	21.6	10.1	2.4

Note: The rolling sample is 1975Q3 to 1997Q3 (ninety periods). At each date in the sample, eighty observations are used to estimate the VAR. The forecasts are computed based on the optimal values of λ and ι (the weight of the prior in the BVAR with Minnesota priors) chosen ex ante.

DSGE prior increases the forecasting performance relative to that of an unrestricted VAR. All numbers are positive and most of them are large, indicating a substantial improvement in forecast accuracy.

Since unrestricted VARs are often overparameterized, they are seldom used in practice for forecasting because of the imprecision with which they are estimated. The results in Table 1 are interesting because they show that the restrictions coming from the DSGE model can alleviate this problem. However, from these results one still does not know whether DSGE-VAR can be relied upon as a forecasting tool. Hence, the second question we ask in this section is, How does the accuracy of the forecasts from DSGE-VAR compare with that of benchmark forecasting models?

Table 2 addresses this question. The benchmark chosen here is a VAR with a Minnesota prior, a standard one in the forecasting literature. The Minnesota prior shrinks the parameter estimates of the VAR toward a unit root in levels (or logarithmic levels).[12] Unlike Table 1, Table 2 has both positive and negative numbers, indicating that the VAR with a Minnesota prior is a tougher competitor than the unrestricted VAR. For federal funds rate forecasts, the VAR with Minnesota prior has the upper hand. However, for both inflation and output growth, DSGE-VAR generally outperforms the BVAR with a Minnesota prior in terms of forecasting accuracy, and the gain generally increases with the forecast horizon.

This section has shown that the DSGE-VAR forecasts can be regarded as competitive relative to a standard benchmark, particularly for inflation but also for real output growth. Of course, it would be interesting to know how DSGE-VAR fares relative to other benchmarks, such as FRB/US or commercial models. At this stage, however, the comparison might be premature, as these models are based on dozens of variables while in the current application the DSGE-VAR includes only three. In future research we plan to apply the DSGE-VAR procedure to a more sophisticated DSGE model, such as, for instance, the one in Christiano, Eichenbaum, and Evans (2001). The resulting application would then include enough variables to make the comparison with FRB/US or commercial models meaningful.

Policy Experiments with DSGE-VAR

For DSGE-VAR to be a useful tool for policy analysis, being competitive in terms of forecasting is not enough. DSGE-VAR needs to be able to address policy questions such as the following: (1) What would be the impact on real output growth and inflation of a 50 basis point cut in the federal funds rate? (2) What would be the impact on the volatility of real output growth and inflation, and ultimately on people's welfare, of changing the policy rule followed by the Federal Reserve?[13]

Models that can address the first type of questions are called "identified" in the literature. They are so named because they are able to identify the impact (impulse-response) of monetary policy shocks, as distinguished from other disturbances in the economy, and therefore assess the consequence of a shock that moves the federal funds rate down by 50 basis points. DSGE models are clearly identified. To see what happens after a 50 basis point shock to the variables of interest, one simply feeds a monetary policy shock that generates a 50 point drop in the federal funds rate into the model. Cowles Commission–style models are also identified to the extent that they contain an equation describing monetary policy. As far as VARs are concerned, the papers by Bernanke (1986) and Sims (1986) show how to obtain such identification. Sims and Zha (1998) extend this framework to BVARs, that is, to VARs with priors. The next section discusses identification in the context of DSGE-VAR.

The second question is different in nature from the first one. The monetary policy shock of the first question can be seen as a one-time disturbance that would not affect the view that market participants have of the Fed. The shift in the policy rule of the second question is likely to affect the view of market participants and their expectations. Because of the Lucas critique, the set of models that can successfully address the second question grows thinner relative to those that address the first question. This is not to say that Cowles Commission–style models and VARs cannot successfully address *any* policy-shift type of question.[14] However, there are some regime shifts that these models may not be able to address. DSGE-VAR addresses regime shifts, trying to strike a balance between the forecasting accuracy of BVARs and the compliance to the Lucas critique of DSGE models. An example of the resulting procedure is discussed later in the article.

Identification

To understand how identification works in the DSGE-VAR procedure, one may find it helpful to review the identification problem in standard VARs (see Hamilton 1994, chap. 11). The problem is as follows: One can easily estimate the variance-covariance matrix of the VAR innovations U in equation (1), which we called Σ_u. The problem is that these innovations do not have an economic interpretation: They are not shocks to monetary policy, technology, or government spending, etc. One would like to have a mapping—call it Ω—between these economically interpretable shocks, which we call E, and the shocks that we cannot interpret, U:

(9) $U = \Omega E.$

With Ω in hand, it is straightforward to compute impulse responses to, say, monetary policy shocks, which are one of the elements of E. Using equation (9), one can feed monetary policy shocks into U and then use equation (1) to feed the U shock into the variables of interest, the Ys. The identification problem is that Ω cannot in general be recovered from the data.[15] Identified VARs address this problem by imposing restrictions (zero restrictions, sign restrictions, etc.) on the matrix Ω. The approach taken here is to learn about Ω from the DSGE model at hand, consistent with the rest of the procedure. Once we learn about Ω, the impulse responses are obtained from equation (1) using $\Phi^{\lambda}(\theta)_{DSGE\text{-}VAR}$ as the parameter estimate.

For the sake of simplicity, we do not delve into the technical details of the identification procedure (see Del Negro and Schorfheide, forthcoming, section 4.3). A comment about the role of λ in the identification procedure is in order, however. As λ increases, DSGE-VAR will tend to coincide with the VAR approximation of the DSGE model. Hence, the impulse responses from DSGE-VAR will become closer and closer to those from the DSGE model. Figure 1 makes this point visually. The figure plots the impulse responses of (cumulative) real output growth, inflation, and the federal funds rate to a monetary policy shock. The impulse responses are based on the sample 1981Q4–2001Q3. The gold lines in the plot are the impulse responses for the DSGE model, the solid black lines are the mean impulse responses for DSGE-VAR, and the dashed lines are 90 percent confidence bands, which measure the uncertainty surrounding the estimates for the impulse responses. One can readily see that as λ increases from 0.5 to 5, the mean impulse responses for DSGE-VAR move closer to the DSGE model's impulse responses, and the bands narrow.

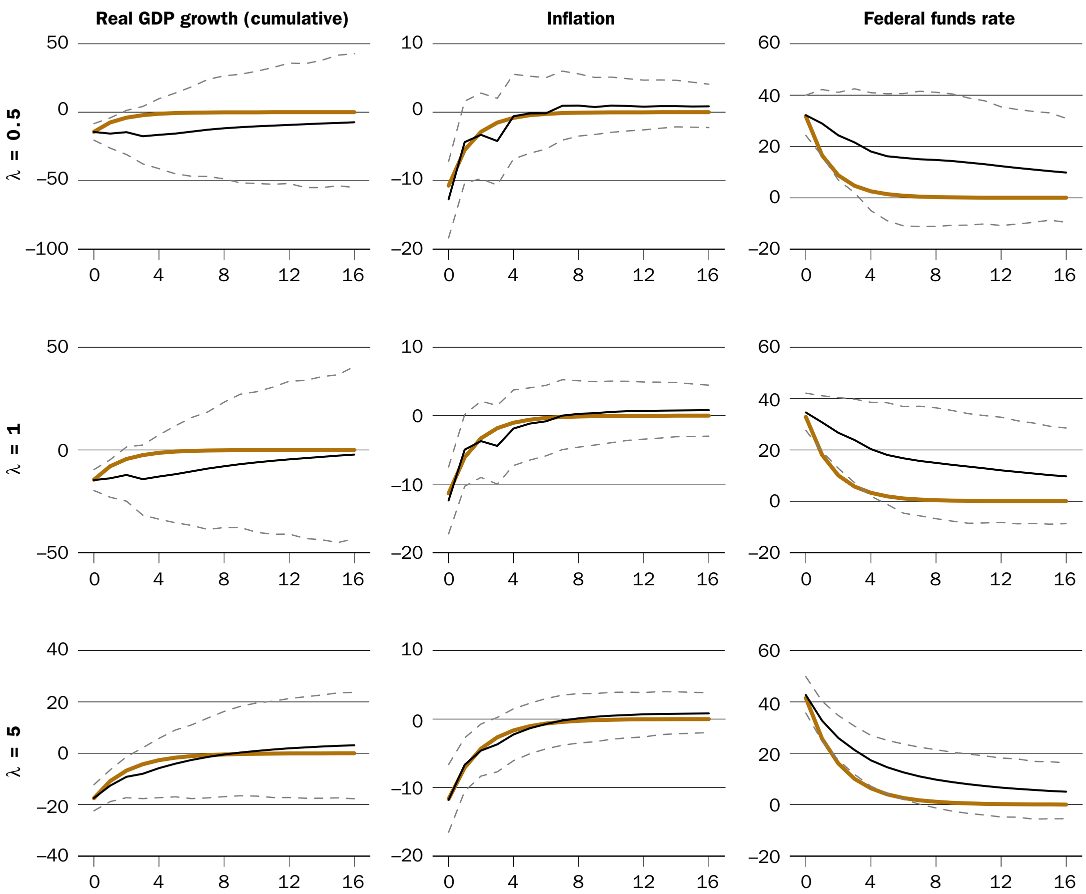
In the best of all possible worlds we would have a DSGE model that forecasts well, so we could forget the VAR correction.

In our procedure λ is computed endogenously and measures the extent to which we can trust the DSGE model used as a prior. We therefore view positively the fact that the identification procedure

12. Since two of the variables, real output and the price level, enter the VAR as growth rates, in the equations corresponding to these variables we shrink the coefficient on the first lag of the "own" variable toward zero. For instance, in the real output equation we shrink the coefficient on the first lag of real output growth toward zero. This restriction corresponds to the unit root in log level. In the equation corresponding to the federal funds rate, the prior on the first lag is one since this variable enters as a level.
13. In asking this question we assume that the Federal Reserve implicitly follows a monetary policy rule, as in Taylor (1993). Whether this is indeed the case is an issue beyond the scope of this paper.
14. See Leeper and Zha (2003) for an interesting analysis of what identified VARs can and cannot address.
15. Note that since $var(U) = \Sigma_u$, it must be the case that $\Omega var(E)\Omega' = \Sigma_u$. Because of this restriction, it is customary to decompose Ω as $\Omega = chol(\Sigma_u)\Omega^*$, where $chol(\Sigma_u)$ is the Cholesky decomposition of Σ_u, Ω^* is an orthonormal matrix, and $var(E)$ is the identity.

FIGURE 1

Impulse Response Functions to Monetary Policy Shocks



Notes: The solid black lines represent the posterior means of the VAR impulse response functions. The dashed lines are 90 percent confidence bands. The gold lines represent the mean impulse responses from the DSGE model. The impulse responses are based on the sample 1983Q3 to 2003Q2.

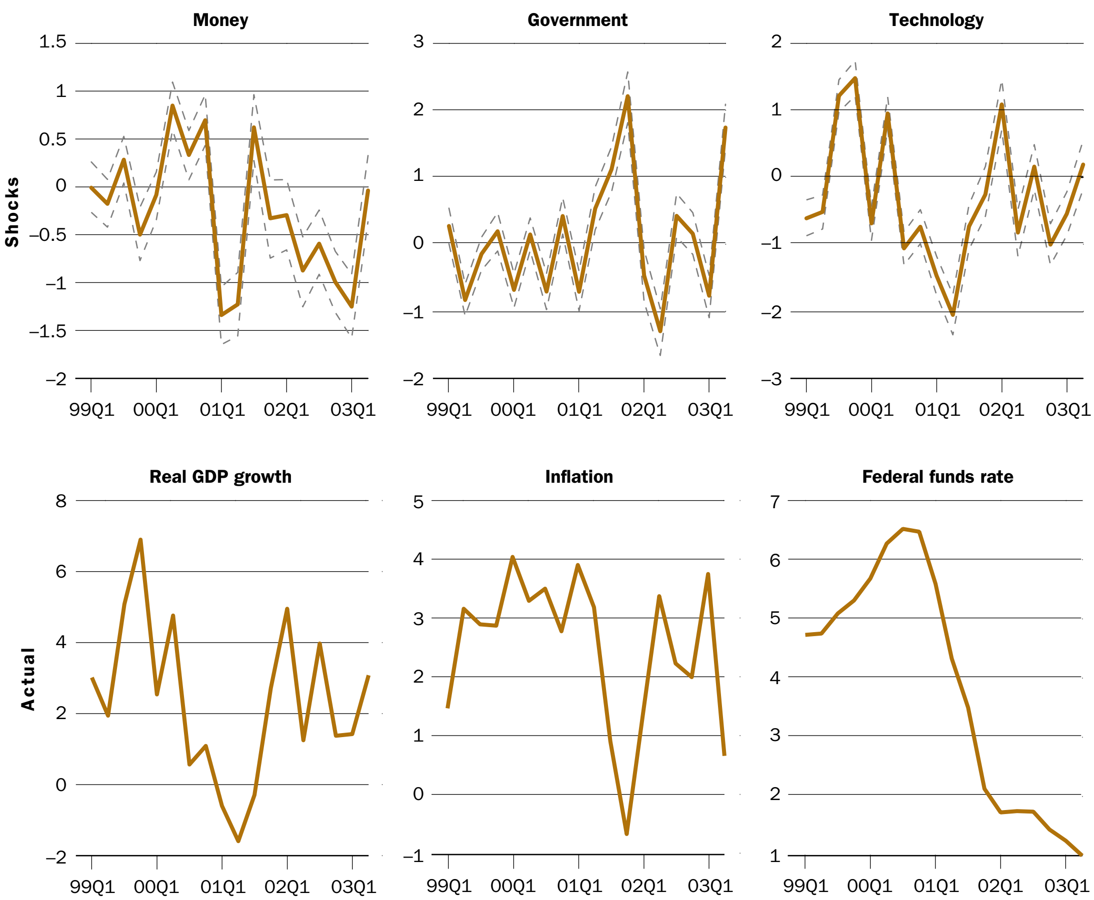
hinges on λ: The higher λ, the more we feel confident about the DSGE model at hand and the more reasonable it becomes to use it as a base for identification. Our approach therefore complements the existing literature, where economic theory is often used as an implicit metric to decide whether a given identification procedure works or not.

We now use the identification in DSGE-VAR to address an issue that is relevant to the current policy discussion: What shocks hit the economy during the past four years and, in particular, during the recession? Was the recession the result of monetary policy shocks, as some have claimed, or was it the result of technology or other shocks? Figure 2 plots the time paths of the identified shocks—that is, the *E* variables in equation (9)—as well as the actual paths of the variables entering the VAR: real output growth, inflation, and the federal funds rate.

As described in the model section, the identified shocks are (1) monetary policy shocks, ε_{Rt}; (2) government spending shocks, g_t; and (3) technology shocks, z_t. To describe the findings in Figure 2, it is necessary to discuss the impulse response functions with respect to these shocks, plotted in Figure 3. Impulse response functions simply trace the impact of a one-standard-deviation shock on the variables of interest. A one-standard-deviation shock can be interpreted as the average shock. The impulse responses in Figure 3 are obtained for a value of λ equal to 1, which is the same value under which the identified shocks in Figure 2 are obtained. Although the impulse responses change with λ, as shown in Figure 1, the overall conclusions of this exercise are fairly robust to the choice of λ. As in Figure 1, the gold lines in the plot are the impulse responses for the DSGE model, the solid black lines are the mean

FIGURE 2

Real GDP Growth, Inflation, the Federal Funds Rate, and Identified Shocks from the DSGE-VAR



Notes: The solid lines in the three upper plots represent the posterior means of the identified shocks from the DSGE-VAR (1999Q2–2003Q2). The dashed lines are 90 percent confidence bands. The sold lines in the three lower plots represent the actual paths of real GDP growth, inflation, and the federal funds rate. The estimates are based on the sample 1983Q3 to 2003Q2.

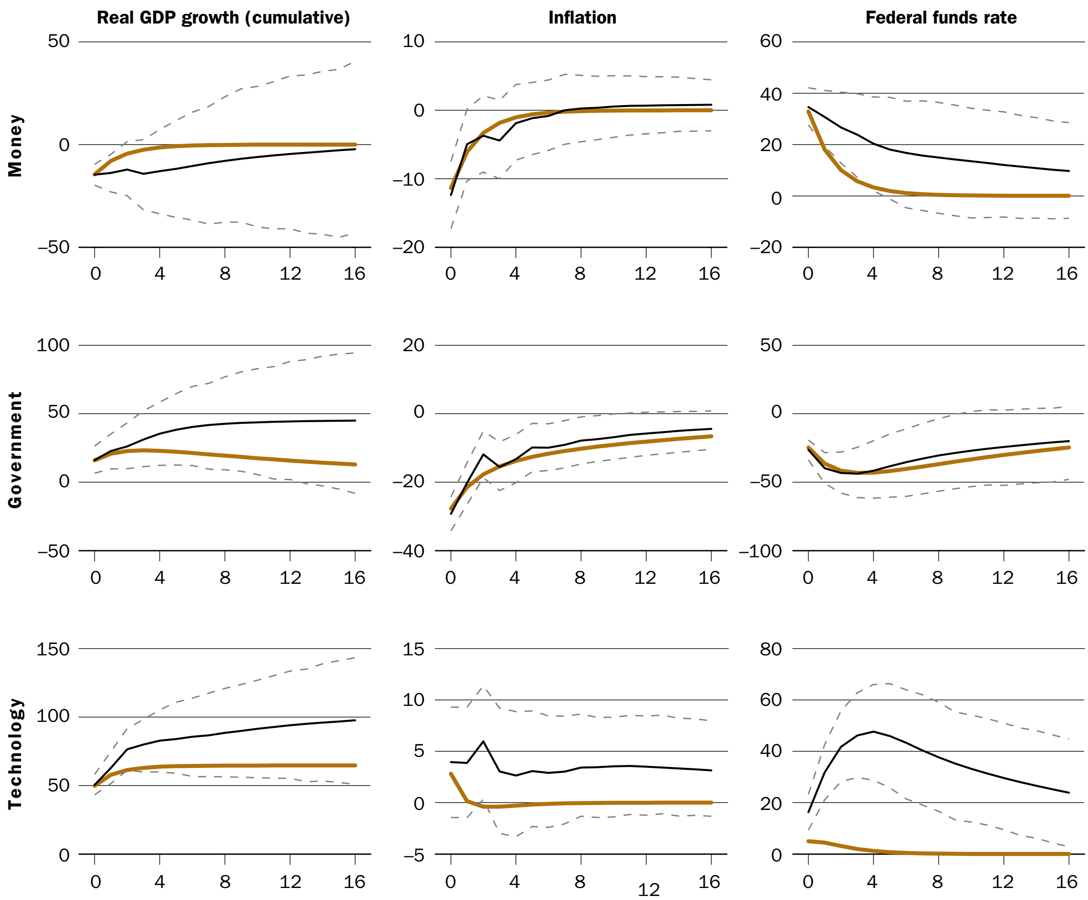
impulse responses for DSGE-VAR, and the dashed lines are 90 percent confidence bands.

The impulse responses of output, inflation, and interest rates to a monetary policy shock conform to well-known patterns. A one-standard-deviation shock raises the federal funds rate by 25–30 basis points, decreases real output growth, and lowers inflation (by convention, here a positive monetary policy shock is contractionary). Note that the response of output in DSGE-VAR is more persistent than in the model. The impulse responses to a government shock deserve more explanation because what is called a government shock in the model is not what people generally have in mind. In the model, a positive government shock is essentially equivalent to a shock to the marginal utility of consumption: For given output, an increase in government spending reduces the resources available from consumption and hence increases the marginal utility of consumption.

The increase in the marginal utility of consumption has two effects. On the cost side, it lowers the real wage and hence the marginal cost faced by firms because in equilibrium the real wage is inversely proportional to the marginal utility of consumption: Since agents value their wages more, all else being equal, they need to be paid less. In sticky price models, a decrease in the marginal cost paid by the firm has the effect of lowering inflation. This reasoning explains the negative impact of a government shock on inflation. On the supply side, the other effect of an increase in the marginal utility of consumption is an increase in output: Again, since agents value output more, they have an incentive to produce more. This reasoning explains the positive impact of a government shock on the output growth

FIGURE 3

Impulse Response Functions to All Shocks



Notes: The solid black lines represent the posterior means of the DSGE-VAR impulse response functions for $\lambda = 1$. The dashed lines are 90 percent confidence bands. The gold lines represent the mean impulse responses from the DSGE model. The impulse responses are based on the sample 1983Q3 to 2003Q2.

rate, which is in any case fairly small. The response of the federal funds rate simply mirrors the decline in inflation as it feeds through the Taylor rule. In summary, positive (negative) government spending shocks in the model look very much like positive (negative) oil price shocks, which drive inflation down (up) and output up (down). Finally, technology shocks drive output up. Since the technology shocks in the model are permanent, the increase in output is permanent as well. The impact on inflation is negligible and insignificant.

The shocks plotted in the three upper panels of Figure 2 are measured in terms of standard deviations: A value of 1 (–1) indicates a positive (negative) shock of one standard deviation. In interpreting the plots, one must bear in mind that shocks between –1 and 1 are the norm while shocks outside this range are the exception. The path of monetary policy shocks indicates that by and large such shocks were not responsible for the last recession. It is true that before the recession most monetary policy shocks were positive (contractionary), but they were fairly small. After the start of the recession, most monetary policy shocks were negative, indicating an accommodative monetary policy stance. In particular, according to the model the beginning of 2001 witnessed two large expansionary shocks.

The driving forces of the recession, according to the model, were technology shocks. Figure 2 shows that technology shocks were positive in 1999 but then turned negative, and sizably so, in 2000 and 2001. The only large positive technology shock was associated with the output rebound in the first quarter of 2002. Finally, government spending shocks were negligible up to the third quarter of 2001, when a large positive shock occurred, associated with the sharp decline in inflation. As Figure 2 shows, government spending shocks have the largest impact on inflation. The recent decline in inflation, resulting from the decline in energy prices, is also associated with a positive government spending shock. This result is not surprising because the effect of government spending shocks in the model is similar to the perceived effect of oil shocks in reality, as discussed above. As energy shocks are not part of the model, their effect is likely attributed to government spending shocks. This remark underscores that the analysis just conducted is in many ways heroic because it is done with a very stylized model and using only a few variables. Yet the purpose of the analysis was to illustrate how, in general, DSGE-VAR can be used to uncover the disturbances affecting the economy.

Regime Shifts

This section describes how DSGE-VAR works under a hypothetical policy experiment. Let us put ourselves in the shoes of Paul Volcker as he took office as chairman of the Federal Reserve Board at the end of the second quarter of 1979. Suppose that he had two options: of being either soft on inflation (labeled policy A) or tough on inflation (labeled policy B). In terms of the Taylor rule in equation (8), policy A corresponds to a low reaction to deviations of inflation from target in the Taylor rule (a low value for ψ_1, say, $\psi_1 = 1.1$) while policy B corresponds to a high value for ψ_1 (say, $\psi_1 = 1.7$).[16] In this hypothetical policy experiment, Chairman Volcker uses macroeconomic stability—measured by the standard deviations of output growth, inflation, and the interest rate in the next twenty years—as the criterion to choose between policies A and B.

To understand how the policy experiment under DSGE-VAR works, it is instructive to see how it would work under a DSGE model. Recall that θ is the vector of deep parameters and that ψ_1 is one of the elements of this vector. Let us assume that the only difference between policy A and B lies in the choice of ψ_1. To perform the policy experiment under a DSGE model, one would estimate the remaining elements of θ using pre-1979Q3 data. Call θ_p, $p = A, B$ the vector of deep parameters corresponding to policies A and B. One would then use the DSGE model to make twenty-year forecasts for the variables of interest and, finally, compute the standard deviation of the forecast paths. To the extent that the dynamics of the DSGE model are reasonably well approximated by a VAR, the forecasts can be obtained from a vector autoregression with coefficients:

$$(10)\quad \Phi^{\lambda=\infty}(\theta_p)_{DSGE\text{-}VAR} = \Gamma^*_{xx}(\theta_p)^{-1}\Gamma^*_{xy}(\theta_p), \quad \text{for } p = A, B.$$

Note that in equation (10) the second moments $\Gamma^*_{xx}(\theta_p)$ and $\Gamma^*_{xy}(\theta_p)$ are computed in full compliance with the Lucas critique. That is, these second moments reflect the fact that agents would behave differently when policy moves from A to B.

To perform the policy experiment under the DSGE-VAR procedure, one would estimate the vector of deep parameters θ using pre-1979Q3 data as described earlier in the article. One would then replace the estimate of ψ_1 with the values 1.1 for policy A and 1.7 for policy B and obtain twenty-year forecasts for the variables of interest using equation (5), which is shown below written in a slightly different way:

$$\Phi^{\lambda}(\theta_p)_{DSGE\text{-}VAR} = \left(\Gamma^*_{xx}(\theta_p) + \frac{1}{\lambda}\frac{X'X}{T}\right)^{-1} \left(\Gamma^*_{xy}(\theta_p) + \frac{1}{\lambda}\frac{X'Y}{T}\right),$$

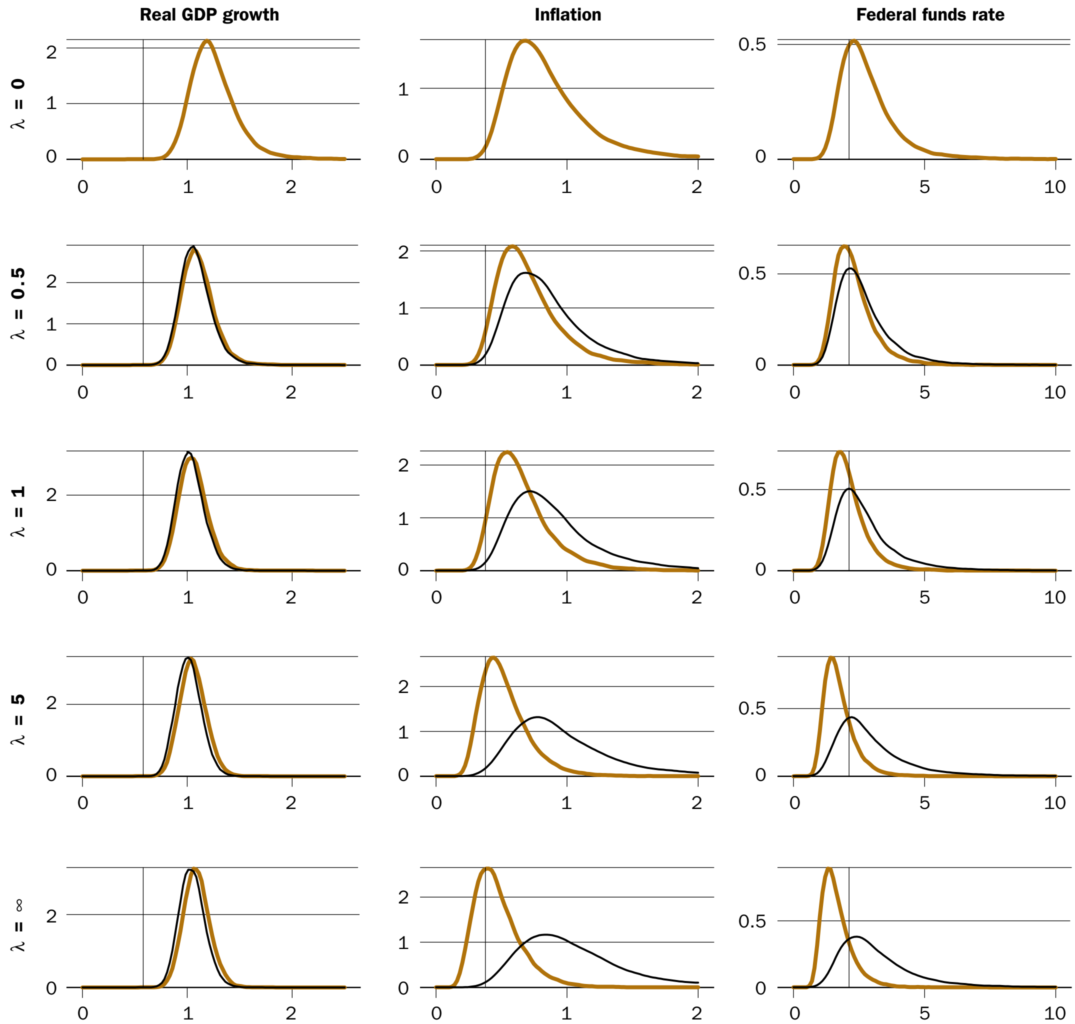
$p = A, B$, where Y and X represent the pre-1979Q3 data. Note that the second moments computed from the data, $(X'X)/T$ and $(X'Y)/T$, do not depend on policy and do not reflect the change in the agent's behavior resulting from the policy shift. This outcome implies that policy experiments under the DSGE-VAR procedure are in full compliance with the Lucas critique only in the $\lambda = \infty$ case. For λ less than infinity, the backward-looking components $(X'X)/T$ and $(X'Y)/T$ are still present. To the extent that the DSGE model does not fit the data well enough, these terms work as a data-driven "correction" to achieve a good forecasting performance. Clearly, in the best of all possible worlds we would have a DSGE model that forecasts well, so we could set $\lambda = \infty$ and forget the backward-looking correction.

The remainder of the section shows the results of the Volcker policy experiment. Figure 4 plots the distributions of outcomes according to policies A and B. Since the assumed criterion of choice is macroeconomic stability, the outcomes we are inter-

16. The values of ψ_1 used in the two policy regimes are broadly consistent with estimated Taylor-rule inflation coefficients obtained over pre- and post-Volcker sample periods by authors such as Clarida, Gali, and Gertler (2000).

FIGURE 4

Effects of a Policy Regime Shift



Notes: The vertical lines correspond to the sample standard deviation of the actual data from 1982Q4 to 1999Q2. The solid black and gold lines are posterior predictive distributions of sample standard deviations for the same time period, obtained using data up to 1979Q2. The solid black line corresponds to $\psi_1 = 1.1$; the gold line corresponds to $\psi_1 = 1.7$.

ested in are the standard deviations of the variables of interest. Remember that for each policy option there is not only one possible outcome but a whole distribution of outcomes, reflecting the uncertainty about the parameters of the model as well the shocks that may hit the economy. Figure 4 is organized as a matrix. The columns of the matrix correspond to real output growth, inflation, and the interest rate, respectively. The rows of the matrix correspond to the relative weight of artificial versus actual data in the augmented sample. The first row ($\lambda = 0$) uses only actual data: This amounts to using the unrestricted VAR only. The last row ($\lambda = \infty$) uses the DSGE model only. The rows in between show the results for values of λ that are between 0 and ∞. Each entry of the matrix plots the standard deviation of the corresponding macroeconomic variable according to option A (solid black line) and option B (gold line). For each plot the vertical line shows what actually occurred from 1982Q4 to 1999Q2.[17]

Notice first that when the DSGE model is not used (first row, $\lambda = 0$) the black and gold lines overlap since the effect of the policy change is embodied only in the artificial data (the $\Gamma^*_{xx}(\theta_p)$ and $\Gamma^*_{xy}(\theta_p)$ terms), which have no weight in this case. As the weight of the artificial data (λ) increases, the predictions from policy A and policy B start to diverge. The forecasts suggest that policy B (tough on inflation) delivers lower variability in both inflation and the interest rate than does policy A (soft on inflation): The gold densities are shifted to the left of, and are narrower than, the black densities. Interestingly, policy B delivers not only a lower variability in inflation, as expected, but also a lower variability of interest rates in spite of the fact that the interest rate reacts more, and not less, to inflation under policy B. This effect works through agents' expectations: Since agents expect monetary policy to reign in inflation under policy B, they will expect lower inflation variability. Their expectations will be realized, and in equilibrium the interest rate will not have to move much. In other words, the threat to react to inflation is enough to lower inflation variability, avoiding wide swings in interest rates.

Although this is a one-time experiment and not a test of the forecasting accuracy of the model, it is interesting to consider how accurate the predictions from DSGE-VAR are in this case. Again, for each plot the dotted vertical lines correspond to the sample standard deviation of the actual data from 1982Q4 to 1999Q2. As far as output is concerned, there is no difference across policies. This result is expected because the difference between policy A and policy B regards the response to inflation and not to output. Not surprisingly, both models overpredict the standard deviation of real output growth: Both the parameters of the BVAR and those of the DSGE model are estimated using data up to 1979, that is, a period in which real output volatility was higher than in the 1980s and 1990s. In terms of inflation, policy B is clearly more on target than policy A is, as it should be since the Taylor rule parameters in policy B are broadly consistent with estimated coefficients obtained over the post-Volcker sample period. Policy A overpredicts the variability of inflation. Also, its forecasts are much more uncertain than those from policy B. For interest rate prediction, for high values of λ policy B appears to underpredict the volatility of the federal funds rate. Policy A, on the other hand, tends to overpredict the rate's volatility. As discussed earlier, the current application of DSGE-VAR is not very accurate in forecasting interest rates.

Summary

This article describes the workings of DSGE-VAR, a procedure that aims to combine VARs and DSGE models. The ultimate goal of the procedure is to provide a proper assessment of the impact of different monetary policy rules and at the same time provide a tool that can also be relied upon for forecasting. It may well be that in the not-too-distant future a full-fledged DSGE model will attain both goals. In the meanwhile, DSGE-VAR may provide a viable alternative to the models that are currently in use for forecasting and policy analysis.

17. Following Clarida, Gali, and Gertler (2000), we compute the actual excluding the pre-1983 disinflation period.

REFERENCES

Altug, Sumru. 1989. Time-to-build and aggregate fluctuations: Some new evidence. *International Economic Review* 30, no. 4:889–920.

Bernanke, Ben S. 1986. Alternative explanations of the money-income correlation. *Carnegie-Rochester Series on Public Policy* 25 (Autumn): 49–99.

Brayton, Flint, Andrew Levin, Ralph Tryon, and John C. Williams. 1997. The evolution of macro models at the Federal Reserve Board. *Carnegie-Rochester Conference Series on Public Policy* 47 (December): 43–81.

Brayton, Flint, Eileen Mauskopf, Dave Reifschneider, Peter Tinsley, and John Williams. 1997. The role of expectations in the FRB/US macroeconomic model. *Federal Reserve Bulletin* 83 (April): 227–45.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 2001. Nominal rigidities and the dynamic effects of a shock to monetary policy. Federal Reserve Bank of Chicago Working Paper 2001-08.

Clarida, Richard, Jordi Gali, and Mark Gertler. 2000. Monetary policy rules and macroeconomic stability: Evidence and some theory. *Quarterly Journal of Economics* 115, no. 1:147–80.

Del Negro, Marco, and Frank Schorfheide. Forthcoming. Priors from general equilibrium models for VARs. *International Economic Review*.

Diebold, Francis X. 1998. The past, present, and future of macroeconomic forecasting. *Journal of Economic Perspectives* 12 (Spring): 175–92.

Doan, Thomas, Robert Litterman, and Christopher A. Sims. 1984. Forecasting and conditional projections using realistic prior distributions. *Econometric Reviews* 3, no. 1:1–100.

Fair, Ray. 1992. The Cowles Commission approach, real business cycle theories, and new Keynesian economics. Yale University, Cowles Foundation Discussion Paper No. 1004.

Hamilton, James D. 1994. *Time series analysis*. Princeton, N.J.: Princeton University Press.

Hodrick, Robert J., and Edward C. Prescott. 1997. Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit, and Banking* 29, no. 1:1–16.

Ingram, Beth F., and Charles H. Whiteman. 1994. Supplanting the "Minnesota" prior: Forecasting macroeconomic time series using real business cycle model priors. *Journal of Monetary Economics* 34 (December): 497–510.

Ireland, Peter N. 1997. A small, structural, quarterly model for monetary policy evaluation. *Carnegie-Rochester Conference Series on Public Policy* 47 (December): 83–108.

Kydland, Finn E., and Edward C. Prescott. 1982. Time to build and aggregate fluctuations. *Econometrica* 50 (November): 1345–70.

———. 1996. The computational experiment: An econometric tool. *Journal of Economic Perspectives* 10 (Winter): 69–85.

Leeper, Eric M., and Christopher A. Sims. 1994. Toward a modern macroeconomic model usable for policy analysis. *NBER macroeconomics annual*. Cambridge and London: MIT Press.

Leeper, Eric M., Christopher A. Sims, and Tao Zha. 1996. What does monetary policy do? *Brookings Papers on Economic Activity* 2:1–78.

Leeper, Eric M., and Tao Zha. 2003. Modest policy interventions. *Journal of Monetary Economics* 50 (November): 1673–1700.

Litterman, Robert B. 1986. Forecasting with Bayesian vector autoregressions—five years of experience. *Journal of Business and Economic Statistics* 4 (January): 25–38.

Lucas, Robert E. 1976. Econometric policy evaluation: A critique. In *The Phillips curve and the labor market*, edited by Karl Brunner and Allan Meltzer. *Carnegie-Rochester Conference Series* 1. Amsterdam: North-Holland.

Reifschneider, David L., David J. Stockton, and David W. Wilcox. 1997. Econometric models and the monetary policy process. *Carnegie-Rochester Conference Series on Public Policy* 47 (December): 1–37.

Robertson, John C., and Ellis W. Tallman. 1999. Vector autoregressions: Forecasting and reality. Federal Reserve Bank of Atlanta *Economic Review* 84 (First Quarter): 4–18.

Sargent, Thomas J. 1984. Autoregressions, expectations, and advice. *American Economic Review* 74, no. 2:408–15.

Schorfheide, Frank. 2000. Loss function-based evaluation of DSGE models. *Journal of Applied Econometrics* 15 (November–December): 645–70.

Sims, Christopher A. 1980. Macroeconomics and reality. *Econometrica* 48 (January): 1–48.

———. 1982. Policy analysis with econometric models. *Brookings Papers on Economic Activity* 1:107–64.

———. 1986. Are forecasting models usable for policy analysis? Federal Reserve Bank of Minneapolis *Quarterly Review* 10 (Winter): 2–16.

———. 2002. The role of models and probabilities in the monetary policy process. *Brooking Papers on Economic Activity* 2:1–40.

Sims, Christopher A., and Tao Zha. 1998. Bayesian methods for dynamic multivariate models. *International Economic Review* 39, no. 4:949–68.

Smith, Anthony A., Jr. 1993. Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics* 8:S63–S84.

Taylor, John B. 1993. Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy* 39:195–214.

Theil, Henry, and Arthur S. Goldberger. 1961. On pure and mixed estimation in economics. *International Economic Review* 2:65–78.

Zha, Tao. 1998. A dynamic multivariate model for use in formulating policy. Federal Reserve Bank of Atlanta *Economic Review* 83 (First Quarter): 16–29.