Accounting for Corporate Behavior

John A. Weinberg

The year 2002 was one of great tumult for the American corporation. As the year began, news of accounting irregularities at energy giant Enron was unfolding at a rapid pace. These revelations would ultimately lead to the demise of that firm and its auditor Arthur Andersen. But Enron was not an isolated case, as other accounting scandals soon followed at WorldCom and Global Crossing in the telecommunications industry and at other prominent companies in different sectors. In July of 2002, Forbes.com published a "corporate scandal sheet" listing some twenty companies that were under investigation by the Securities and Exchange Commission (SEC) or other government authority.[1] Of these cases, the vast majority involved misreporting of corporate earnings.

These allegations certainly created the appearance of a general phenomenon in corporate finance, and the resulting loss of confidence in financial reporting practices arguably contributed to the weakness of markets for corporate securities. The fact that many of the problems were surfacing in industries that had been at the center of the new economy euphoria of the late 1990s contributed to the sense of malaise by shaking investor confidence in the economy's fundamental prospects. In most of the recent cases, the discovery of accounting improprieties was accompanied by a spectacular decline of high-flying stocks and, in a number of cases, criminal charges against corporate executives. Consequently, the state of corporate governance and accounting became the dominant business news story of the year.

To some observers, the recent events confirm a sense that the stock market boom of the 1990s was artificial—a "bubble" backed solely by unrealistic

This article first appeared in the Bank's 2002 Annual Report. It benefited from conversations with a number of the author's colleagues in the Research Department and from careful and critical readings by Tom Humphrey, Jeff Lacker, Ned Prescott, John Walter, and Alice Felmlee. The views expressed herein are the author's and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] Patsuris (2002).

expectations with no grounding in economic fundamentals. According to this view, investors' bloated expectations were nourished by the fictitious performance results reported by some firms. In the aftermath of these events, Congress enacted a new law known as the Sarbanes-Oxley Act to reform corporate accounting practices and the corporate governance tools that are intended to ensure sound financial reporting.

The attention received by the various scandals and the legislative response might easily create the impression that a fundamental flaw developed in the American system of corporate governance and finance during the late 1990s. It *does* appear that the sheer number of cases in which companies have been forced to make significant restatements of their accounts, largely as the result of SEC action, has risen in recent years. Beginning in 1998 with large earnings restatements by such companies as Sunbeam and Waste Management and with a heightened commitment by the SEC, under then chairman Arthur Levitt, to police misleading statements of earnings, the number of cases rose significantly above the dozen or so per year that was common in the 1980s.[2] While the frequency and magnitude of recent cases seem to be greater than in the past, accounting scandals are not new. Episodes of fraudulent accounting have occurred repeatedly in the history of U.S. financial markets.

In the aftermath of the stock market crash of 1929, public attention and congressional investigation led to allegations of unsavory practices by some financial market participants during the preceding boom. This activity led directly to the creation of the Securities and Exchange Commission in 1934. One of the founding principles of this agency was that "companies publicly offering securities. . . must tell the public the truth about their businesses."[3] The creation of the SEC, however, did not eliminate the problem, and scandals associated with dubious accounting remained a feature of the financial landscape. In 1987 a number of associations for accounting and finance professionals organized a National Commission on Fraudulent Financial Reporting. The commission studied cases from the 1980s and characterized the typical case as involving a relatively small company with weak internal controls. Although incidents of fraud were often triggered by a financial strain or sudden downturn in a company's real performance, the companies involved were usually from industries that had been experiencing relatively rapid growth. So while the size of companies involved in recent cases may be atypical, the occurrence of scandals in high-growth firms fits the established pattern.

Does fraudulent financial reporting represent the Achilles' heel of U.S. corporate finance? This essay addresses such questions by examining the

[2] Alternative means of tallying the number of cases are found in Richardson et al. (2002) and Financial Executives Research Foundation Inc. (2001). By both measures, there was a marked increase in the number of cases in the late 1990s.

[3] From the SEC Web page.

problem of financial reporting in the context of the fundamental problem of corporate governance. Broadly stated, that fundamental problem is the need for a large group of corporate outsiders (shareholders) to be able to control the incentives of a small group of corporate insiders (management). At the heart of this problem lies a basic and inescapable asymmetry: insiders are much better informed about the opportunities and performance of a business than are any outsiders. This asymmetry presents a challenge that the modern corporation seeks to address in the mechanisms it uses to measure performance and reward managers.

While the tools of corporate governance can limit the effects of the incentive problem inherent in the corporate form, they cannot eliminate it. Ultimately, there are times when shareholders just have to trust that management is acting in their best interest and realize that their trust will sometimes be violated. Still, management has a powerful interest in earning and preserving the trust of investors. With trust comes an enhanced willingness of investors to provide funds, resulting in reduced funding costs for the business. That is, the behavior of corporate insiders is disciplined by their desire or need to raise funds in financial markets. This discipline favors efficient corporate governance arrangements.

As discussed in the next section, there are a variety of tools that a corporation might use to control managerial discretion, ranging from the makeup and role of the board of directors to the firm's relationship with its external auditor. To say that such tools are applied efficiently is to say that managers will adopt a tool as long as its benefit outweighs its cost. In the absence of government intervention, the forces of competition among self-interested market participants (both insiders and outsiders) will tend to lead to an efficient set of governance tools. It bears repeating, though, that these tools do not eliminate the fundamental problem of corporate governance. The observation of apparent failures, such as the accounting scandals of 2002, is not inconsistent, however, with a generally well-functioning market for corporate finance. Still, such episodes often provoke a political response, as occurred during the Great Depression and again in 2002 with the Sarbanes-Oxley Act. Through these interventions, the government has assumed a role in managing the relationship between shareholders and management.

The final sections of the essay consider the role of a government authority in setting and enforcing rules. After reviewing the functions of the SEC, discussion turns to the Sarbanes-Oxley Act, the provisions of which can be classified into two broad categories. Parts of the act attempt to improve corporate behavior by mandating certain aspects of the design of the audit committee or the relationship between the firm and its external auditor. The discussion in this essay suggests that there is reason to doubt that such provisions, by themselves, can do much to reduce fraud. Other parts of the act deal more with enforcement and the penalties for infractions. These provisions are more

likely to have a direct effect on incentives. An open question is whether this effect is desirable. Since reducing fraud is costly, it is unlikely that reducing it to zero would be cost effective from society's point of view. Further, it is unrealistic to expect the new law to bring about a substantial reduction in instances of fraud without an increase in the resources allocated to enforcement. Given that it is in the interest of corporate stakeholders to devise mechanisms that respond efficiently to the fundamental problem of corporate governance, one might doubt that the gains from government intervention will be worth the costs necessary to bring about significant changes in behavior.

1. THE NATURE OF THE MODERN CORPORATION

In the modern American corporation, ownership is typically spread widely over many individuals and institutions. As a result, owners as a group cannot effectively manage a business, a task that would require significant coordination and consensus-building. Instead, owners delegate management responsibilities to a hired professional. To be sure, professional managers usually hold some equity in the firms they run. Still, it is common for a manager's ownership stake to be small relative both to the company's total outstanding equity and to the manager's own total wealth.[4]

This description of the modern corporation featuring a separation between widely dispersed ownership and professional management is typically associated with the work of Adolf Berle and Gardiner Means. In their landmark study, *The Modern Corporation and Private Property,* Berle and Means identified the emerging corporate form as a cause for concern. For them, the separation of ownership and control heralded the rise of a managerial class, wielding great economic power but answerable only to itself. Large numbers of widely dispersed shareholders could not possibly exert effective control over management. Berle and Means' main concern was the growing concentration of economic power in a few hands and the coincident decline in the competitiveness of markets. At the heart of this problem was what they saw as the impossibility of absentee owners disciplining management.

Without adequate control by shareholders in the Berle and Means view, managers would be free to pursue endeavors that serve their own interests at shareholders' expense. Such actions might include making investments and acquisitions whose main effect would be to expand management's "empire." Managers might also use company resources to provide themselves with desirable perks, such as large and luxurious corporate facilities. These actions

[4] Holderness et al. (1999) present evidence of rising managerial ownership over time. They find that executives and directors, *as a group,* owned an average of 21 percent of the outstanding stock in corporations they ran in 1995, compared to 13 percent in 1935.

could result in the destruction of shareholder wealth and an overall decline in efficiency in the allocation of productive resources.

The experience of the last seventy years and the work of a number of writers on the law and economics of corporate governance have suggested that the modern corporation is perhaps not as ominous a development as imagined by Berle and Means. A field of financial economics has developed that studies the mechanisms available to shareholders for exerting some influence over management's decisions.[5] These tools represent the response of governance arrangements to the forces of supply and demand. That is, managers implement a governance mechanism when they perceive that its benefits exceed its costs. The use of these tools, however, cannot eliminate the fundamental asymmetry between managers and owners. Even under the best possible arrangement, corporate insiders will be better informed than outsiders.

The most obvious mechanism for affecting an executive's behavior is the compensation arrangement between the firm and the executive. This tool, however, is also the most subject to problems arising from the separation of ownership and control. Just as it would be difficult for owners to coordinate in directly running the firm, so it is difficult for them to coordinate employment contract negotiations with managers. In practice, this task falls to the board of directors, who, while intended to represent owners, are often essentially controlled by management. In terms of this relationship, management can benefit by creating a strong and independent board. This move signals to owners that management is seeking to constrain its own discretion. Ultimately, however, shareholders face the same challenge in assessing the board's independence as they do in evaluating management's behavior. The close contact the board has with management makes its independence hard to guarantee.

Another source of control available to owners comes from the legal protections provided by corporate law. Shareholders can bring lawsuits against management for certain types of misbehavior, including fraud and self-dealing, by which a manager unjustly enriches himself through transactions with the firm. Loans from the corporation to an executive at preferential interest rates can be an example of self-dealing. Of course use of the courts to discipline management also requires coordination among the widespread group of shareholders. In such cases, coordination can be facilitated by class-action lawsuits, where a number of shareholders come together as the plaintiff. Beyond suing management for specific actions of fraud or theft, however, shareholders' legal rights are limited by a general presumption in the law that management is best positioned to take actions in the firm's best business interest.[6] For instance, if management chooses between two possible investment projects, dissatisfied shareholders would find it very difficult to make a case that management's

[5] Shleifer and Vishny (1997) provide a survey of this literature.

[6] This point is emphasized by Roe (2002).

choice was driven by self-interest as opposed to shareholder value. So, while legal recourse can be an important tool for policing certain types of managerial malfeasance, such recourse cannot serve to constrain the broad discretion that management enjoys in running the business.

Notice that this discussion of tools for controlling managers' behavior has referred repeatedly to the coordination problem facing widely dispersed shareholders. Clearly, the severity of this problem depends on the degree of dispersion. The more concentrated the ownership, the more likely it is that large shareholders will take an active role in negotiating contracts and monitoring the behavior of management. Concentrated ownership comes at a cost, though. For an investor to hold a large share of a large firm requires a substantial commitment of wealth without the benefits of risk diversification. Alternatively, many investors can pool their funds into institutions that own large blocks of stock in corporations. This arrangement does not solve the corporate governance problem of controlling incentives; however, it simply shifts the problem to that of governing the shareholding institutions.

In spite of the burden it places on shareholders, concentrated ownership has won favor as an approach to corporate governance in some settings. In some developed economies, banks hold large shares of equity in firms and also participate more actively in their governance than do financial institutions in the United States. In this country, leveraged buyouts emerged in the 1980s as a technique for taking over companies. In a leveraged buyout, ownership becomes concentrated as an individual or group acquires the firm's equity, financed through the issuance of debt. Some see the leveraged buyout wave as a means of forcing businesses to dispose of excess capacity or reverse unsuccessful acquisitions.[7] In most cases, these transactions resulted in a temporary concentration of ownership, since subsequent sales of equity eventually led back to more dispersed ownership. It seems that, at least in the legal and financial environment of the United States, the benefits of diversification associated with less concentrated ownership are great enough to make firms and their shareholders willing to face the related governance challenges.[8] Still, there is considerable variation in the concentration of ownership among large U.S. corporations, leading some observers to conclude that this feature of modern corporations responds to the relative costs and benefits.[9]

A leveraged buyout is a special type of takeover, an additional tool for controlling managers' incentives. If a firm is badly managed, another firm

[7] Holmstrom and Kaplan (2001) discuss the role of the leveraged buyouts of the 1980s in aligning managerial and shareholder interests.

[8] Roe (1994) argues that ownership concentration in the United States has been constrained by a variety of legal restrictions. While this argument might temper one's conclusion that the benefits of dispersed ownership outweigh the costs, the leveraged buyout episode provides an example of concentration that was consistent with the legal environment and yet did not last.

[9] Demsetz and Lehn (1985) make this argument.

can acquire it, installing new management and improving its use of resources so as to increase profits. The market for corporate control, the market in which mergers and acquisitions take place, serves two purposes in corporate governance.[10] First, as just noted, it is sometimes the easiest means by which ineffective managers can be replaced. Second, the threat of replacement can help give managers an incentive to behave well. Takeovers, however, can be costly transactions and may not be worth the effort unless the potential improvement in a firm's performance is substantial.

The threat of a takeover introduces the idea that a manager's current behavior could bring about personal costs in the future. Similarly, a manager may have an interest in building and maintaining a reputation for effectively serving shareholders' interest. Such a reputation could enhance the manager's set of future professional opportunities. While reputation can be a powerful incentive device, like other tools, it is not perfect. There will always be *some* circumstances in which a manager will find it in his best interest to take advantage of his good reputation for a short-run gain, even though he realizes that his reputation will suffer in the long run. For example, a manager might "milk" his reputation by issuing misleading reports on the company's performance in order to meet targets needed for additional compensation.

The imperfections of reputation as a disciplining tool are due to the nature of the corporate governance problem and the relationship between ownership and management. Any tools shareholders have to control management's incentives are limited by a basic informational advantage that management enjoys. Because management has superior information about the firm's opportunities, prospects, and performance, shareholders can never be perfectly certain in their evaluation of management's actions and behavior.

2. CORPORATE GOVERNANCE AS AN AGENCY PROBLEM

At the heart of issues related to corporate governance lies what economists call an agency (or principal-agent) problem. Such a problem often arises when two parties enter into a contractual relationship, like that of employer-employee or borrower-lender. The defining characteristic of an agency problem is that one party, the principal, cannot directly control or prescribe the actions of the other party, the agent. Usually, this lack of control results from the agent having superior information about the endeavor that is of mutual interest to both parties. In the employer-employee relationship, this information gap is often related to the completion of daily tasks. Unable to monitor all of their employees' habits, bosses base workers' salaries on performance to induce those

[10] Henry Manne (1965) was an early advocate of the beneficial incentive effect on the market for corporate control.

workers to put appropriate effort into their work.[11] Another common example of an agency problem includes insurance relationships. In auto insurance, for instance, the insurer cannot directly monitor the car owner's driving habits, which directly affect the probability of a claim being filed. Typical features of insurance contracts such as deductibles serve to enhance the owner's incentive to exercise care.

In interpreting corporate governance as an agency problem, it is common to identify top corporate management as the agent and owners as the principal. While both management and ownership are typically composed of a number of individuals, the basic tensions that arise in an agency relationship can be seen quite clearly if one thinks of each of the opposing parties as a single individual. In this hypothetical relationship, an owner (the principal) hires a manager (the agent) to run a business. The owner is not actively involved in the affairs of the firm and, therefore, is not as well-informed as the manager about the opportunities available to the firm. Also, it may not be practical for the owner to monitor the manager's every action. Accordingly, the control that the owner exerts over the manager is primarily indirect. Since the owner can expect the manager to take actions that maximize his own return, the owner can try to structure the compensation policy so that the manager does well when the business does well. This policy could be supplemented by a mutual understanding of conditions under which the manager's employment might be terminated.

The agency perspective is certainly consistent with a significant part of compensation for corporate executives being contingent on firm performance. Equity grants to executives and equity options are common examples of performance-based compensation. Besides direct compensation, principals have a number of other tools available to affect agents' incentives. As discussed earlier, the tools available to shareholders include termination of top executives' employment, the possibility of a hostile takeover, and the right to sue executive management for certain types of misbehavior. Like direct compensation policy, all of these tools involve consequences for management that depend on corporate performance. Hence, the effective use of such tools requires that principals be able to assess agents' performance.

In the usual formulation of an agency problem, the agent takes an action that affects the business's profits, and the principal pays the agent an amount that depends on the level of those profits. This procedure presumes that the principal is able to assess the firm's profits. But the very same features of a modern corporation that make it difficult for principals (shareholders) to monitor actions taken by agents (corporate management) also create an asymmetry

[11] Classic treatments of agency problems are given by Holmstrom (1979) for the general analysis of moral hazard and Jensen and Meckling (1976) for the characterization of corporate governance as an agency problem.

in the ability of shareholders and managers to track the firm's performance. Since owners cannot directly observe all of the firm's expenses and sales revenues, they must rely to some extent on the manager's reports about such measures of performance. As discussed in the next section, the problem of corporate governance is a compound agency problem: shareholders suffer from both an inability to directly control management's actions and an inability to easily obtain information necessary to assess management's performance.

The characterization of corporate governance as an agency problem might lead one to doubt the ability of market forces to achieve efficient outcomes in this setting. But an agency problem is not a source of market failure. Rather, agents' and principals' unequal access to relevant information is simply a condition of the economic environment. In this environment, participants will evaluate contractual arrangements taking into account the effects on the incentives for all parties involved. An individual or a firm that can devise a contract with improved incentive effects will have an advantage in attracting other participants. In this way, market forces will tend to lead to efficient contracts. Accordingly, the economic view of corporate governance is that firms will seek executive compensation policies and other governance mechanisms that provide the best possible incentive for management to work in shareholders' best interest. The ultimate governance structure chosen does not eliminate the agency problem but is a rational, best response to that problem, balancing the costs and benefits of managerial discretion.

3. ACCOUNTING FOR CORPORATE PERFORMANCE

All of the tools intended to influence the incentives and behavior of managers require that outsiders be able to assess when the firm is performing well and when it is performing poorly. If the manager's compensation is tied to the corporation's stock price, then investors, whose behavior determines the stock price, must be able to make inferences about the firm's true performance and prospects from the information available. If management's discipline comes from the threat of a takeover, then potential acquirers must also be able to make such assessments.

The challenge for effective market discipline (whether in the capital market or in the market for corporate control) is in getting information held by corporate insiders out into the open. As a general matter, insiders have an interest in providing the market with reliable information. If by doing so they can reduce the uncertainty associated with investing in their firm, then they can reduce the firm's cost of capital. But it's not enough for a manager to simply say, "I'm going to release reliable financial information about my business on an annual (or quarterly or other interval) basis." The believability of such a statement is limited because there will always be some circumstances in which a manager can benefit in the short term by not being fully transparent.

The difficulty in securing reliable information may be most apparent when a manager's compensation is directly tied to accounting-based performance measures. Since these measures are generated inside the firm, essentially by the same group of people whose decisions are driving the business's performance, the opportunity for manipulation is present. Certainly, accounting standards set by professional organizations can limit the discretion available to corporate insiders. A great deal of discretion remains, however. The academic accounting literature refers to such manipulation of current performance measures as "earnings management."

An alternative to executive compensation that depends on current performance as reported by the firm is compensation that depends on the market's perception of current performance. That is, compensation can be tied to the behavior of the firm's stock price. In this way, rather than depending on self-reported numbers, executives' rewards depend on investors' collective evaluation of the firm's performance. Compensation schemes based on this type of investor evaluation include plans that award bonuses based on stock price performance as well as those that offer direct grants of equity or equity options to managers.

Unfortunately, tying compensation to stock price performance hardly eliminates a manager's incentive to manipulate accounting numbers. If accounting numbers are generally believed by investors to provide reliable information about a company's performance, then those investors' trading behavior will cause stock prices to respond to accounting reports. This responsiveness could create an incentive for managers to manipulate accounting numbers in order to boost stock prices. Note, however, that if investors viewed earnings management and other forms of accounting manipulation as pervasive, they would tend to ignore reported numbers. In this case, stock prices would be unresponsive to accounting numbers, and managers would have little reason to manipulate reports (although they would also have little incentive to exert any effort or resources to creating accurate reports). The fact that we do observe cases of manipulation suggests that investors do not ignore accounting numbers, as they would if they expected all reports to be misleading. That is, the prevailing environment appears to be one in which serious instances of fraud are occasional rather than pervasive.

In summary, the design of a system of rewards for a corporation's top executives has two conflicting goals. To give executives an incentive to take actions that maximize shareholder value, compensation needs to be sensitive to the firm's performance. But the measurement of performance is subject to manipulation by the firm's management, and the incentive for such manipulation grows with the sensitivity of rewards to measured performance.

This tension limits the ability of compensation plans to effectively manage executives' incentives.[12]

Are there tools that a corporation can use to lessen the possibility of manipulated reporting and thereby improve the incentive structure for corporate executives? One possible tool is an external check on a firm's reported performance. A primary source for this check in public corporations is an external auditor. By becoming familiar with a client and its performance, an auditor can get a sense for the appropriateness of the choices made by the firm in preparing its reports. Of course, every case of fraudulent financial reporting by corporations, including those in the last year, involves the failure of an external auditor to detect or disclose problems. Clearly, an external audit is not a fail-safe protection against misreporting. A significant part of the Sarbanes-Oxley legislation was therefore devoted to improving the incentives of accounting firms in their role as external auditors.

An external audit is limited in its ability to prevent fraudulent reporting. First, many observers argue that an auditor's role is limited to certifying that a client's financial statements were prepared in accordance with professional accounting standards. Making this determination does not automatically enable an auditor to identify fraud. Others counter that an auditor's knowledge of a client's operations makes the auditor better positioned than other outsiders to assess the veracity of the client's reports. In this view, audit effectiveness in deterring fraud is as much a matter of willingness as ability.

One aspect of auditors' incentives that has received a great deal of attention is the degree to which the auditor's interests are independent of the interests of the client's management.[13] Some observers argue that the objectivity of large accounting firms when serving as external auditors is compromised by a desire to gain and retain lucrative consulting relationships with those clients. Even before the events of 2002, momentum was growing for the idea of separating the audit and consulting businesses into separate firms. Although the Sarbanes-Oxley Act did not require such a separation, some audit firms have taken the step of spinning off their consulting businesses. This step, however, does not guarantee auditor independence. Ultimately, an auditor works for its client, and there are always strong market forces driving a service provider to give the client what the client wants. If the client is willing to pay more for an audit that overlooks some questionable numbers than the (expected) costs to the auditor for providing such an audit, then that demand will likely be met. In general, a client's desire to maintain credibility with investors gives it a strong interest in the reliability of the auditor's work. Even so, there will always be

[12] Lacker and Weinberg (1989) analyze an agency problem in which the agent can manipulate the performance measure.

[13] Levitt (2000) discusses this point.

some cases in which a client and an auditor find themselves willing to breach the public's trust for a short-term gain.

Some observers suggest that making the hiring of the auditor the responsibility of a company's board of directors, in particular the board's audit committee, can prevent complicity between management and external auditors. This arrangement is indeed a standard procedure in large corporations. Still, the ability of such an arrangement to enhance auditor independence hinges on the independence of the board and its audit committee. Unfortunately, there appears to be no simple mechanism for ensuring the independence of directors charged with overseeing a firm's audit relationships. In 1987 the National Commission on Fraudulent Financial Reporting found that among the most common characteristics of cases that resulted in enforcement actions by the Securities and Exchange Commission was weak or inactive audit committees or committees that had members with business ties to the firm or its executives. While such characteristics can often be seen clearly after the fact, it can be more difficult and costly for investors or other outsiders to discriminate among firms based on the general quality of their governance arrangements before problems have surfaced. While an outside investor can learn about the members of the audit committee and how often it meets, investors are less able to assess how much care the committee puts into its work.

The difficulty in guaranteeing the release of reliable information arises directly from the fundamental problem of corporate governance. In a business enterprise characterized by a separation of ownership and control, those in control have exclusive access to information that would be useful to the outside owners of the firm. Any outsider that the firm hires to verify that the information it releases is correct becomes, in effect, an insider. Once an auditor, for instance, acquires sufficient knowledge about a client to assess its management's reports, that auditor faces incentive problems analogous to those faced by management. So, while an external audit might be part of the appropriate response to the agency problem between management and investors, an audit also creates a new and analogous agency problem between investors and an auditor.

An alternative approach to monitoring the information released by a firm is for this monitoring to be done by parties that have no contractual relationship with the firm's management. Investors, as a group, would benefit from the increased credibility of accounting numbers this situation would provide. Suppose that a small number of individual investors spent the resources necessary to assess the truthfulness of a firm's report. Those investors could then make trades based on the results of their investigation. In an efficient capital market, the results would then be revealed in the firm's stock price. In this way, the firm's management would suffer the consequences (in the form of a lower stock price) of making misleading reports. The problem with this scenario is that while only a few investors incur the cost of the investigation

and producing the information, all investors receive the benefit. Individual investors will have a limited incentive to incur such costs when other investors can free ride on their efforts. Because it is difficult for dispersed shareholders to coordinate information-gathering efforts, such free riding might occur and is just a further reflection of the fundamental problem of corporate governance.

The free-riding problem that comes when investors produce information about a firm can be reduced if an individual investor owns a large fraction of a firm's shares. As discussed in the second section, however, concentrated ownership has costs and does not necessarily resolve the information and incentive problems inherent in corporate governance. An alternative approach to the free-riding problem, and one that extends beyond the governance arrangements of an individual firm, is the creation of a membership organization that evaluates firms and their reporting behavior. Firms would be willing to pay a fee to join such an organization if membership served as a seal of approval for reporting practices. Members would then enjoy the benefits of reduced funding costs that come with credibility.

One type of membership organization that could contribute to improved financial reporting is a stock exchange. As the next section discusses, the New York Stock Exchange (NYSE) was a leader in establishing disclosure rules prior to the stock market crash of 1929. The political response to the crash was the creation of the Securities and Exchange Commission, which took over some of the responsibilities that might otherwise fall to a private membership organization. Hence, a government body like the SEC might substitute for private arrangements in monitoring corporate accounting behavior. The main source of incentives for a government body is its sensitivity to political sentiments. While political pressure can be an effective source of incentives, its effectiveness can also vary depending on political and economic conditions. If government monitoring replaces some information production by private market participants, it is still possible for such a hybrid system of corporate monitoring to be efficient as long as market participants base their actions on accurate beliefs about the effectiveness of government monitoring.

Given the existence of a governmental entity charged with policing the accounting behavior of public corporations, how much policing should that entity do? Should it carefully investigate every firm's reported numbers? This would be an expensive undertaking. The purpose of this policing activity is to enhance the incentives for corporate managements and their auditors to file accurate reports. At the same time, this goal should be pursued in a cost-effective manner. To do this, there is a second tool, beyond investigation, that the agency can use to affect incentives. The agency can also vary the punishment imposed on firms that are found to have violated the standards of honest reporting. At a minimum, this punishment simply involves the reduction in stock price that occurs when a firm is forced to make a restatement of earnings or other important accounting numbers. This minimum punishment, imposed

entirely by market forces, can be substantial.[14] To toughen punishment, the government authority can impose fines or even criminal penalties.

To increase corporate managers' incentive for truthful accounting, a government authority can either increase resources spent on monitoring firms' reports or increase penalties imposed for discovered infractions. Relying on large penalties allows the authority to economize on monitoring costs but, as long as monitoring is imperfect, raises the likelihood of wrongly penalizing firms. The Sarbanes-Oxley Act has provisions that affect both of these margins of enforcement. The following sections describe enforcement in the United States before and after Sarbanes-Oxley.

4. GOVERNMENT ENFORCEMENT OF CORPORATE HONESTY

Before the creation of the Securities and Exchange Commission in 1934, regulation of disclosures by firms issuing public securities was a state matter. Various states had "blue sky laws," so named because they were intended to "check stock swindlers so barefaced they would sell building lots in the blue sky."[15] These laws, which specified disclosures required of firms seeking to register and issue securities, had limited impact because they did not apply to the issuance of securities across state lines. An issuer could register securities in one state but offer them for sale in other states through the mail. The issuer would then be subject only to the laws of the state in which the securities were registered. The New York Stock Exchange offered an alternative, private form of regulation with listing requirements that were generally more stringent than those in the state laws. The NYSE also encouraged listing firms to make regular, audited reports on their income and financial position. This practice was nearly universal on the New York Stock Exchange by the late 1920s. The many competing exchanges at the time had weaker rules.

One of the key provisions of the Securities Exchange Act of 1934 was a requirement that all firms issuing stock file annual and quarterly reports with the SEC. In general, however, the act did not give finely detailed instructions to the commission. Rather, the SEC was granted the authority to issue rules "where appropriate in the public interest or for the protection of investors."[16] As with many of its powers, the SEC's authority with regard to the treatment of information disclosed by firms was left to an evolutionary process.

In the form into which it has evolved, the SEC reviews financial reports, taking one of a number of possible actions when problems are found. There are two broad classes of filings that the Corporate Finance Division of the

[14] Richardson et al. (2002).
[15] Seligman (1982, 44).
[16] Seligman (1982, 100).

SEC reviews—transactional and periodic filings. Transactional filings contain information relevant to particular transactions, such as the issuance of new securities or mergers and acquisitions. Periodic filings are the annual and quarterly filings, as well as the annual report to shareholders. Among the options available to the Corporate Finance Division if problems are found in a firm's disclosures is to refer the case to the Division of Enforcement.

Given its limited resources, it is impossible for the SEC to review all of the filings that come under its authority. In general, more attention is paid to transactional filings. In particular, all transactional filings go through an initial review, or screening process, to identify those warranting a closer examination. Many periodic filings do not even receive the initial screening. While the agency's goal has been to review every firm's annual 10-K report at least once every three years, it has not had the resources to realize that goal. In 2002 around half of all public companies had not had such a review in the last three years.[17] It is possible that the extraordinary nature of recent scandals has been due in part to the failure of the SEC's enforcement capabilities to keep up with the growth of securities market activity.

5. THE SARBANES-OXLEY ACT OF 2002

In the aftermath of the accounting scandals of 2002, Congress enacted the Sarbanes-Oxley Act, aimed at enhancing corporate responsibility and reforming the practice of corporate accounting. The law contains provisions pertaining to both companies issuing securities and those in the auditing profession. Some parts of the act articulate rules for companies and their auditors, while other parts focus more on enforcement of these rules.[18]

The most prominent provisions dealing with companies that issue securities include obligations for the top executives and rules regarding the audit committee. The act requires the chief executive and financial officers to sign a firm's annual and quarterly filings with the SEC. The signatures will be taken to certify that, to the best of the executives' knowledge, the filings give a fair and honest representation of the firm's financial condition and operating performance. By not fulfilling this signature requirement, executives could face the possibility of significant criminal penalties.

The sections of the act that deal with the audit committee seek to promote the independence of directors serving on that committee. To this end, the act requires that members of the audit committee have no other business relationship with the company. That is, those directors should receive no compensation from the firm other than their director's fee. The act also instructs audit committees to establish formal procedures for handling complaints about

[17] United States Senate, Committee on Governmental Affairs (2002).

[18] A summary of the act is found in Davis and Murray (2002).

accounting matters, whether the complaints come from inside or outside of the firm. Finally, the committee must include a member who is a "financial expert," as defined by the SEC, or explain publicly why it has no such expert.

Like its attempt to promote audit committee independence, the act contains provisions regarding a similar relationship between a firm and its auditor. A number of these provisions are intended to keep the auditor from getting "too close" to the firm. Hence, the act specifies a number of nonaudit services that an accounting firm may not provide to its audit clients. The act also requires audit firms to rotate the lead partner responsible for a client at least once every five years. Further, the act calls on the SEC to study the feasibility of requiring companies to periodically change their audit firm.

With regard to enforcement, the act includes both some new requirements for the SEC in its review of company filings and the creation of a new body, the Public Company Accounting Oversight Board. The PCAOB is intended to be an independent supervisory body for the auditing industry with which all firms performing audits of public companies must register. This board is charged with the task of establishing standards and rules governing the operation of public accounting firms. As put forth in Sarbanes-Oxley, these standards must include a minimum period of time over which audit workpapers must be maintained for possible examination by the PCAOB. Other rules would involve internal controls that audit firms must put in place to protect the quality and integrity of their work.

Sarbanes-Oxley gives the PCAOB the task of inspecting audit firms on a regular basis, with annual inspection required for the largest firms.[19] In addition to examining a firm's compliance with rules regarding organization and internal controls, inspections may include reviews of specific audit engagements. The PCAOB may impose penalties that include fines as well as the termination of an audit firm's registration. Such termination would imply a firm's exit from the audit business.

In addition to creating the new board to supervise the audit industry, the act gives the SEC greater responsibilities in reviewing disclosures by public companies. The act spells out factors that the SEC should use in prioritizing its reviews. For instance, firms that have issued material restatements of financial results or those whose stock prices have experienced significant volatility should receive priority treatment. Further, Sarbanes-Oxley requires that no company be reviewed less than once every three years. Other sections of the act that deal with enforcement prescribe penalties for specific abuses and extend the statute of limitations for private securities fraud litigation.

The goal of the Sarbanes-Oxley Act is to alter the incentives of corporate managements and their auditors so as to reduce the frequency of fraudulent

[19] Firms preparing audit reports for more than one hundred companies per year will be inspected annually.

financial reporting. In evaluating the act, one can take this goal as given and try to assess the act's likely impact on actual behavior of market participants. Alternatively, one could focus on the goal itself. The act is presumably based on the belief that we currently have too much fraud in corporate disclosures. But what is the right amount of fraud? Total elimination of fraud, if even feasible, is unlikely to be economically desirable. As argued earlier, reducing fraud is costly. It requires the expenditure of resources by some party to evaluate the public statements of companies and a further resource cost to impose consequences on those firms determined to have made false reports. Reduction in fraud is only economically efficient or desirable as long as the incremental costs of enforcement are less than the social gain from improved financial reporting.

What are the social benefits from improved credibility of corporate information? A reduction in the perceived likelihood of fraud brings with it similar benefits to other risk reductions perceived by investors. For example, investors become more willing to provide funds to corporations that issue public securities, resulting in a reduction in the cost of capital for those firms. Other things being equal, improved credibility should also lead to more investment by public companies and an overall expansion of the corporate sector. Again, however, any such gain must be weighed against the corresponding costs.

Is there any reason to believe that a private market for corporate finance, without any government intervention, would not result in an efficient level of corporate honesty? Economic theory suggests that the answer is no. It is true that the production of information necessary to discover fraud has some characteristics of a public good. For example, many people stand to benefit from an individual's efforts in investigating a company. While public goods can impede the efficiency of private market outcomes, the benefits of information production accrue to a well-defined group of market participants in this case. Companies subject to heightened investigative scrutiny enjoy lower costs of capital.

In principle, one can imagine this type of investigative activity being undertaken by a private membership organization. Companies that join would voluntarily subject their accounting reports to close review. Failure to comply with the organization's standards could be punished with expulsion. This organization could fund its activities through membership fees paid by the participating companies. It would only attract members if the benefits of membership, in the form of reduced costs of capital, exceeded the cost of membership. That is, such an organization would be successful if it could improve at low cost the credibility of its members' reported information. Still, even if successful, the organization would most likely not eliminate the potential for fraud among its members. There would always be some circumstances in which the short-run gain from reporting false numbers would outweigh the risk of discovery and expulsion.

Before the stock market crash of 1929, the New York Stock Exchange was operating in some ways much like the hypothetical organization just described. Investigations after the crash, which uncovered instances of misleading or fraudulent reporting by issuers of securities, found relatively fewer abuses among companies issuing stock on the NYSE.[20] One might reasonably conjecture that through such institutions the U.S. financial markets would have evolved into an efficient set of arrangements for promoting corporate honesty. While consideration of this possibility would make an interesting intellectual exercise, it is not what happened. Instead, as often occurs in American politics, Congress responded to a crisis with the creation of a government entity. In this case, a government entity charged with policing the behavior of companies that issue public securities. The presence of such an agency might well dilute private market participants' incentives to engage in such policing activities. If so, then reliance on the government substitutes for reliance on private arrangements.

Have the SEC's enforcement activities resulted in an efficient level of corporate honesty? This is a difficult determination to make. It is true that known cases of misreporting rose steadily in the 1980s and 1990s and that the events of 2002 represented unprecedented levels of both the number and the size of companies involved. It is also true that over the last two decades, as activity in securities markets grew at a very rapid pace, growth in the SEC's budget lagged, limiting the resources available for the review of corporate reports. In this sense, one might argue that the level of enforcement fell during this period. Whether the current level of enforcement is efficient or not, the Sarbanes-Oxley Act expresses Congress's interest in seeing heightened enforcement so as to reduce the frequency of fraudulent reports.

How effective is Sarbanes-Oxley likely to be in changing the incentives of corporations and their auditors? Many of the act's provisions set rules and standards for ways in which firms should behave or how they should organize themselves and their relationships with auditors. There is reason to be skeptical about the likely effectiveness of these provisions by themselves. These portions of the act mandate that certain things be done inside an issuing firm, for instance, in the organization of the audit committee. But because these actions and organizational changes take place inside the firm, they are subject to the same information problems as all corporate behavior. It is inherently difficult for outsiders, whether market participants or government agencies, to know what goes on inside the firm. The monitoring required to gain this information is costly, and it is unlikely that mandates for changed behavior will have much effect without an increase in the allocation of resources for such monitoring of corporate actions, relationships, and reports.

[20] Seligman (1982, 46).

Other parts of the act appear to call for this increase in the allocation of resources for monitoring activities, both by the SEC and by the newly created PCAOB. Together with the act's provisions concerning penalties, these portions should have a real effect on incentives and behavior. Further, to the extent that these agencies monitor firms' adherence to the general rules and standards specified in the act, monitoring will give force to those provisions. If the goal of the act is to reduce the likelihood of events like Enron and WorldCom, however, monitoring might best be applied to the actual review of corporate reports and accounting firms' audit engagements. Ultimately, such direct review of firms' reports and audit workpapers is the activity that identifies misbehavior. Uncovering and punishing misbehavior is, in turn, the most certain means of altering incentives.

Incentives for deceptive accounting will never be eliminated, and even a firm that follows all of the formal rules in the Sarbanes-Oxley Act will find a way to be deceptive if the expected payoff is big enough. Among the things done by the SEC and PCAOB, the payoff to deception is most effectively limited by the allocation of resources to direct review of reported performance and by bringing penalties to bear where appropriate. Any hope that a real change in corporate behavior can be attained without incurring the costs of paying closer attention to the actual reporting behavior of firms will likely lead to disappointment. Corporate discipline, whether from market forces or government intervention, arises when people outside of the firm incur the costs necessary to learn some of what insiders know.

REFERENCES

Berle, Adolf, and Gardiner Means. 1932. *The Modern Corporation and Private Property.* New York: Commerce Clearing House.

Davis, Harry S., and Megan E. Murray. 2002. "Corporate Responsibility and Accounting Reform." *Banking and Financial Services Policy Report* 21 (November): 1–8.

Demsetz, Harold, and Kenneth Lehn. 1985. "The Structure of Corporate Ownership: Causes and Consequences." *Journal of Political Economy* 93 (December): 1155–77.

Financial Executives Research Foundation Inc. 2001. "Quantitative Measures of the Quality of Financial Reporting" (7 June).

Holderness, Clifford G., Randall S. Krozner, and Dennis P. Sheehan. 1999. "Were the Good Old Days That Good? Changes in Managerial Stock

Ownership Since the Great Depression." *Journal of Finance* 54 (April): 435–69.

Holmstrom, Bengt. 1979. "Moral Hazard and Observability." *Bell Journal of Economics* 10 (Spring): 74–91.

_____, and Steven N. Kaplan. 2001. "Corporate Governance and Merger Activity in the United States: Making Sense of the 1980s and 1990s." *Journal of Economic Perspectives* 15 (Spring): 121–44.

Jensen, Michael C., and William H. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics* 3 (October): 305–60.

Lacker, Jeffrey M., and John A. Weinberg. 1989. "Optimal Contracts Under Costly State Falsification." *Journal of Political Economy* 97 (December): 1345–63.

Levitt, Arthur. 2000. "A Profession at the Crossroads." Speech delivered at the National Association of State Boards of Accountancy, Boston, Mass., 18 September.

Manne, Henry G. 1965. "Mergers and the Market for Corporate Control." *Journal of Political Economy* 73 (April): 110–20.

Patsuris, Penelope. 2002. "The Corporate Scandal Sheet." Forbes.com (25 July).

Richardson, Scott, Irem Tuna, and Min Wu. 2002. "Predicting Earnings Management: The Case of Earnings Restatements." University of Pennsylvania Working Paper (October).

Roe, Mark J. 1994. *Strong Managers, Weak Owners: The Political Roots of American Corporate Finance*. Princeton, N.J.: Princeton University Press.

_____. 2002. "Corporate Law's Limits." *Journal of Legal Studies* 31 (June): 233–71.

Seligman, Joel. 1982. *The Transformation of Wall Street: A History of the Securities and Exchange Commission and Modern Corporate Finance*. Boston: Houghton Mifflin.

Shleifer, Andrei, and Robert W. Vishny. 1997. "A Survey of Corporate Governance." *Journal of Finance* 52 (June): 737–83.

United States Senate, Committee on Governmental Affairs. 2002. "Financial Oversight of Enron: The SEC and Private-Sector Watchdogs." Staff report (8 October).

Japanese Monetary Policy and Deflation

Robert L. Hetzel

Japan is experiencing deflation. Its price level (measured by the GDP deflator) fell about 10 percent from the end of 1997 to the end of 2002. The Bank of Japan (BoJ) possesses the power to end deflation and restore price stability by creating money. To do so, the BoJ needs to adopt a policy of active reserves creation where reserves creation depends upon misses of a target either for money growth or for the price level. With its present policy of demand-driven reserves creation, the BoJ limits reserves creation to the amount of reserves demanded by banks. The high level of reserves held by banks does not indicate an aggressive BoJ policy of reserves provision. The BoJ has only accommodated the increased demand for excess reserves by banks produced by a zero short-term interest rate.

The sole focus of political pressures on the composition of the BoJ's asset portfolio, in particular, on the purchase of nontraditional assets such as stocks and long-term government bonds (JGBs), is misplaced. In the absence of a strategy that makes the amount of bank reserves vary to control money and prices—for example, to eliminate misses in a target for the price level—the acquisition of such assets is comparable to sterilized foreign exchange intervention. Their purchase affects the composition of the public's asset portfolio without increasing bank reserves and money in a way that forces the portfolio rebalancing that stimulates expenditure.

According to popular commentary, monetary policy is impotent to stop deflation. One argument made is that the transmission mechanism linking central bank reserves creation to money and credit creation has been severed. Lacking opportunities to lend, banks hold whatever liquidity the central bank provides as excess reserves. Another argument is that at a zero interest rate a

The views in this paper are solely those of the author, not the Federal Reserve Bank of Richmond or the Federal Reserve System. The author appreciates research assistance from John Hejkal and assistance obtaining data from Kanou Adachi, Toshitaka Sekine, and Takashi Kodama. Margarida Duarte, Milton Friedman, Marvin Goodfriend, Motoo Haruta, and Alexander Wolman provided helpful criticism.

limitless demand for money (a liquidity trap) causes the public to absorb any increase in money rather than spend it.[1]

I dispute these arguments below. Even with zero short-term interest rates, the BoJ can control money creation. Money creation combined with the considerable stability of money demand in Japan will stimulate expenditure. However, to do so, the BoJ must abandon its current policy of market-determined reserves creation that limits reserves to amounts demanded by banks.

In discussing Japanese monetary policy, newspapers make statements like the BoJ's "arsenal of traditional tools [has been] rendered largely ineffective" (*New York Times,* 9 April 2003). Such misperceptions arise from a lack of understanding of basic principles of central banking. For this reason, I review these principles.

In "The Nature of a Central Bank" (Section 1), I explain that a central bank is a creator of money, not a financial intermediary. In "How a Central Bank Controls the Money Stock" (Section 1), I explain money stock determination when the central bank uses an interest rate instrument. Even with an interest rate instrument, central bank control over expenditure derives from its control over reserves creation. When short-term interest rates become zero, the central bank should shift to a strategy of explicit reserves targeting to retain control over expenditure. The BoJ has not made that transition.

Section 1 continues with an explanation of money stock determination with a reserves instrument. With a zero short-term interest rate, the aggregate the central bank must control becomes the monetary base plus government securities yielding zero interest. Section 2 reviews the current BoJ operating strategy. Sections 3 and 4 examine the behavior of money demand. Sections 5 and 6 discuss strategies for ending deflation. Section 7 deals with issues of political economy, and Section 8 argues that current monetary policy procedures leave the Japanese economy unable to adjust to adverse shocks.

---

[1] The current debate over Japanese monetary policy replays the old debate over whether the Federal Reserve System had the power to end deflation in the Great Depression. Milton Friedman (1956, 17) stated the quantity theory view challenging arguments of central bank impotence:

> The quantity theorist…holds that there are important factors affecting the supply of money that do not affect the demand for money.…The classical version of the objection under this head to the quantity theory is the so-called real-bills doctrine: that changes in the demand for money call forth corresponding changes in supply and that supply cannot change otherwise.…
>
> The attack on the quantity theory associated with the Keynesian underemployment analysis is based primarily on an assertion about the [demand for money]. The demand for money, it is said, is infinitely elastic at a "small" positive interest rate. At this interest rate…changes in the real supply of money…have no effect on anything. This is the famous "liquidity trap."

1. HOW THE BOJ CAN CONTROL MONEY CREATION AND YEN EXPENDITURE

An understanding of how a central bank controls money begins with an understanding of the nature of a central bank.

The Nature of a Central Bank

A central bank is not a commercial bank. It creates money rather than intermediates between savers and investors. This distinction is critical because popular commentary dwells on the supposed responsibility of the BoJ to control financial intermediation rather than money creation. Such commentary leads to the misplaced conclusion that the BoJ should concentrate on the structural reform of the financial system. For example, Koll (*Asian Wall Street Journal,* 26 February 2003) turns monetary theory on its head:

> By giving bankers a free ride, the BoJ's zero-rate policy is a root cause of Japan's fundamental problems—excess capacity, excess debt and excess employment. And by preventing a market-based destruction of excess capacity, the BoJ's zero-rate policy has significantly contributed to Japan's deflationary problem. . . . The key task is to raise interest rates. . . . Excess capacity would quickly be cut back.

Commercial banks are financial intermediaries. They acquire the debt of businesses, consumers, and government by issuing their own debt (deposits). Banks make loans and issue deposits up to the point where the marginal return from lending equals the marginal cost of borrowing. They create a broad market for their own debt (deposits) by making it liquid through a guarantee of the par (dollar, yen) value of their deposits. They also provide payment services through the transfer of ownership of deposits.[2] Although banks bundle their intermediation and payment services, they are conceptually distinct.[3]

Because a commercial bank acquires assets until the marginal return of lending equals the marginal cost of issuing liabilities, the marketplace limits the amount of liabilities an individual commercial bank creates. Extension of this logic to a central bank is the essence of the real bills fallacy that the market

[2] As Goodfriend (1990) explains, banks bundle both financial intermediation and payment services because both involve the assessment of credit risk. Because the transfer of ownership of deposits does not occur in real time, but rather involves float or temporary credit extension between institutions, the provision of payments services involves credit evaluation.

[3] During the Depression, several economists (Henry Simons, Lauchlin Currie, and Irving Fisher) advocated 100 percent reserves requirements. That is, the only assets that banks could hold were currency. Banks would provide only payment services. Other financial institutions would issue debt to provide for financial intermediation.

limits central bank asset acquisition and base money creation. However, no such market mechanism exists to limit the issuance of central bank liabilities.

The liabilities of a central bank constitute base money (currency and deposits held with it by commercial banks). The central bank controls base money through its asset acquisition. It can then control money creation and the money price of goods—the price level. The failure to understand this responsibility leads to the belief that the central bank is not responsible for deflation. For example, Miller (*Asian Wall Street Journal,* 28 February 2003) writes, "Deflation . . . is not a monetary problem. It's a problem of the fundamental structure of Japanese industry. . . . The problem of deflation . . . is structural overcapacity."

Another fallacy due to the confusion of a central bank with commercial banks is that a central bank must worry about solvency. Otsuma and Chiba (2003) report that "limits on the central bank's capital make it 'impossible' to expand purchases [of equities]." However, central bank insolvency does not entail the same consequences as for a private corporation. The holders of central bank liabilities cannot run it by turning in currency to the central bank and demanding payment. Commercial banks can ask for currency in place of their deposits with the central bank, but the central bank can simply create additional currency. A change in the market value of a central bank's assets produces no change in the dollar (yen) value of its liabilities. A central bank balance sheet is important not as a measure of solvency but rather as a bookkeeping procedure for keeping track of monetary base creation.

With a positive short-term interest rate, a central bank exerts its control over the money stock through its influence over base money creation. Individuals and banks hold base money to arrange for the finality of payment. The public holds currency to make small transactions. Banks hold reserves to accommodate the public's demand for currency and to clear payments with other banks.

The amount of reserves banks demand to clear deposits varies with the amount of those deposits. Although central banks share money creation with commercial banks, their control over base money creation provides them with control over bank deposits and the money stock. Control over money creation endows central banks with control over the dollar (yen) expenditure of the public. The reason is that money creation induces the public to rebalance its portfolio.

Portfolio Balance

Money is one asset in individuals' portfolios. In order for them to be satisfied with the allocation of their assets, all assets must yield the same return adjusted for risk and liquidity. Equation (1), taken from Friedman (1969b), equates the return between money, government bonds, and capital (a proxy for any

illiquid real asset).[4] The return to money includes the marginal liquidity (nonpecuniary) services yield of money ($MNPS_M$) minus the cost imposed by expected inflation $\left(\frac{1}{P}\frac{dP}{dt}\right)^*$ (or plus the return due to expected deflation). The return to bonds is the marginal liquidity services yield of bonds ($MNPS_B$) plus the explicit interest yield (r_B) and the negative of expected inflation. The marginal real yield on capital is MRY.

$$MNPS_M - \left(\frac{1}{P}\frac{dP}{dt}\right)^* = MNPS_B + r_B - \left(\frac{1}{P}\frac{dP}{dt}\right)^* = MRY. \qquad (1)$$

Purposeful money creation by the central bank not offset by a commensurate price increase causes individuals to rebalance their portfolios. The increase in money lowers the marginal return on money relative to nonmonetary assets by lowering the marginal liquidity services yield on money. When the public attempts to move out of money into nonmonetary assets, it bids up the prices of those assets and lowers their yield. The reduction in yield spurs expenditure.[5]

The fall in yields on nonmonetary assets and the increase in expenditure induce the public to hold a larger real money stock. This equilibrium is temporary because it occurs without a change in the real resources and productive opportunities available to society. Portfolio balance returns only when the price level rises to restore the real money stock to its original value. The limitless ability of a central bank to create money through base money creation allows it to force portfolio rebalancing by the public.[6]

How a Central Bank Controls the Money Stock

Unfortunately, the standard central bank practice of setting a target for the short-term interest rate obscures the fact that central banks control the public's

[4] "[E]ach dollar is…regarded as rendering a variety of services, and the holder of money as altering his money holdings until the value to him of the addition to the total flow of services produced by adding a dollar to his money stock is equal to the reduction in the flow of services produced by subtracting a dollar from each of the other forms in which he holds assets" (Friedman 1956, 14).

[5] "The key feature of this process is that it tends to raise the prices of sources of both producer and consumer services relative to the prices of the services themselves; for example, to raise the prices of houses relative to the rents of dwelling units, or the cost of purchasing a car relative to the cost of renting one. It therefore encourages the production of such sources (this is the stimulus to 'investment'…) and, at the same time, the direct acquisition of services rather than of the source (this is the stimulus to 'consumption' relative to 'savings')" Friedman (1969a, 255–56).

[6] A central bank is not just one among many institutions in the money market influencing credit flows. The way a central bank controls inflation does not depend upon the myriad, ever-changing institutional arrangements that circumscribe financial intermediation. The credit channel emphasized by Bernanke and Gertler (1995) constitutes part of the transmission mechanism of monetary policy; however, it *propagates* monetary shocks.

nominal expenditure through their control of money creation. I will explain this fact through a discussion of money stock determination relevant to interest-rate-targeting procedures. I will also explain how central banks retain control over expenditure with zero short-term interest rates by moving from interest-rate-targeting to reserves-targeting procedures. The discussion will distinguish between indirect control of the monetary base, which occurs with interest rate targeting, and direct control, which occurs with reserves aggregate targeting.

One can understand how a central bank achieves monetary control when it sets an interest rate target by understanding the discipline imposed by such procedures. This discipline is twofold, corresponding to the nominal and real components of an interest rate. A nominal interest rate measures the intertemporal price of a dollar in terms of dollars. A nominal interest rate of 10 percent represents a promise to pay \$1.10 in the future for \$1.00 today. Its real kernel is the real interest rate, which measures the intertemporal price of goods in terms of goods.

The nominal interest rate measures the real interest rate using the monetary (dollar) standard, whose value changes with changes in the price level. The nominal interest rate therefore incorporates an expectation of the change in the price level. Two facts are central: Because the central bank determines the inflation rate, it controls the behavior of this expectation. In contrast, the central bank cannot control the level of the real interest rate in a sustained way.

The real interest rate reflects the pattern of relative scarcity produced by the intertemporal distribution of consumption. A higher value of expected consumption in the future relative to current consumption requires a higher real interest rate. The natural rate of interest, MRY_N, is the real rate of interest in the absence of monetary disturbances. Alternatively, it is the real interest rate yielded by the real business cycle core of an economy with perfectly flexible prices. To control inflation, the central bank must respect the working of the price system by moving its interest rate target, $\bar{r}_B$, in a way that tracks the natural rate.

Consider again formula (1) with MRY set equal to MRY_N and r_B equal to the central bank's interest rate target, $\bar{r}_B$:

$$MNPS_M - \left(\frac{1}{P}\frac{dP}{dt}\right)^* = MNPS_B + \bar{r}_B - \left(\frac{1}{P}\frac{dP}{dt}\right)^* = MRY_N. \quad (2)$$

The central bank must move its interest rate peg, $\bar{r}_B$, in line with movements in the natural rate, MRY_N. For example, if it fails to raise its rate peg $\bar{r}_B$ in line with a rise in the natural rate, it creates base money and money. This money creation makes the marginal nonpecuniary services of money, $MNPS_M$, fall and the first two terms of (1) become less than the last. The public will rebalance its portfolio by moving out of money into illiquid assets.

The resulting rise in the price of illiquid assets stimulates current consumption, and a rise in current consumption relative to expected future consumption restrains the rise in the real rate relative to the rise in the natural rate.[7] However, a central bank does not stockpile the resources necessary to run a commodity stabilization scheme for the real interest rate. If not subsequently offset, the monetary emissions created by transitory divergences between the real rate and the natural rate force changes in the price level.

Central banks perform the ongoing task of tracking the natural rate by raising their rate peg relative to its prevailing value when economic growth strengthens relative to trend, and conversely. A prerequisite for performing this task is to stabilize expected inflation at a value equal to the central bank's inflation target. If the central bank does not tie down the public's expectation of inflation, the behavior of its rate peg becomes a loose cannon.

To summarize, with an interest rate target, to control the nominal money stock in a way that achieves predictable control of the price level, the central bank must fulfill two conditions. Monetary policy must be credible: the public's expectation of inflation must correspond to the central bank's inflation target, $\left[\left(\frac{1}{P}\frac{dP}{dt}\right)^* = \pi^T\right]$. Also, the central bank must vary its rate peg, $\bar{r}_B$, to track changes in the natural rate, that is, to maintain the following equality: $\left[MNPS_B + \bar{r}_B - \left(\frac{1}{P}\frac{dP}{dt}\right)^* = MRY_N\right]$. Given these conditions, the public's trend nominal expenditure growth equals trend real output growth plus trend inflation (equal to the central bank's target).[8]

---

[7] See Goodfriend and King (1997) for a review of optimizing, sticky-price models that deliver this result.

[8] Economists say that with an interest-rate instrument, money is demand determined (at the price set by the central bank). In interpreting this statement, one must remember that the price of money is the price level, not the interest rate. (The goods price of money is the inverse of the price level.) The interest rate is the opportunity cost of holding real money balances. Money is demand determined because the central bank ties down the public's expectation of the future price level.

Current models in the literature with endogenous determination of the price level often omit money as a variable (for example, McCallum 2001). At first pass, this omission seems analogous to a model of the price of pencils that omits the quantity of pencils. However, the central bank's inflation target determines the public's expectation of inflation. That expectation determines the behavior of nominal variables.

With credibility and procedures that provide for tracking changes in the natural rate, the central bank's inflation target controls both money growth and inflation. The central bank's inflation target is the exogenous variable, while money is endogenous. Nevertheless, money remains critical. It is the ability to produce monetary shocks that endows the central bank with control over the public's expectations of inflation.

By assumption in these models, the central bank knows that it controls inflation, sets an inflation target, and pursues a policy consistent with its inflation target. Also, the public knows the target and the policy rule. The behavior of money then offers no independent information about the behavior of prices. That latter assumption becomes questionable in periods when monetary policy changes and the public is slow to learn of the change. For example, for much of the 1960s in the United States, the public formed its expectation of inflation based on prior experience with a commodity standard rather than the actual, inflationary monetary policy. In such a period, the behavior of money predicts inflation.

For the United States in the last two decades, the problem has been inflationary expectations in excess of the Fed's implicit target. On several occasions, as economic growth quickened, these expectations jumped, as measured by the behavior of bond rates. As Goodfriend (1993) documents, the Fed dealt with these "inflation scares" through sharp increases in the funds rate. The ability to contract the monetary base gave the Fed the ability to engineer these funds rate increases.

For Japan, the problem is a "deflation scare." Because short-term interest rates are zero, the BoJ cannot lower interest rates.[9] Instead, it must increase the monetary base directly. The public announcement of a commitment to stabilize the price level accompanied by an expansion of the monetary base could reverse expectations of deflation. However, even if the public continues to expect deflation, monetary base expansion will stimulate expenditure through portfolio rebalancing. A revival of expenditure will eventually make credible a commitment to price stability.

When the central bank controls the monetary base directly by setting a target for a reserves aggregate, the reserves-money multiplier formula highlights the relevant behavioral relationships.[10] Given a reserves target, the money stock becomes a function of the currency-deposits ratio desired by the public and the reserves-deposits ratio desired by the banking system. Banks demand reserves for clearing purposes. With reserves-targeting procedures, the central bank uses the reserves-deposits ratio desired by banks as a lever for controlling the money stock.[11]

With a zero short-term interest rate, the monetary base and short-term Treasury securities become perfect substitutes. In (1), if expected deflation equals the real yield on capital, the short-term interest rate r_B is zero ($MNPS_M = MNSP_B = 0$). Because the marginal liquidity services yield on money then equals zero, the public is sated with liquidity and is indifferent between Treasury securities and money. In this case, the relevant aggregate that the central bank uses to force portfolio rebalancing is the sum of the

[9] Goodfriend (2000) argues that the central bank can make the cost of carry for money positive instead of zero by taxing bank reserves and currency. A negative rather than a zero interest rate then becomes the relevant lower bound.

[10] With interest rate targeting, fluctuations in the reserves-deposits ratio do not affect the money stock because the central bank automatically offsets such fluctuations as a consequence of maintaining its interest rate peg.

The multiple expansion of deposits in response to a reserves injection by the central bank is a textbook construction. With reserves targeting and an interbank market for reserves, a reserves injection by the central bank would produce a reduction in the funds rate relative to the returns that banks earn on assets. Banks would respond by buying assets. The resulting increase in deposits would raise the reserves-deposits ratio. Reserves do not pass sequentially from bank to bank. However, if the interbank rate is zero, a reserves injection could produce the sequence of deposit expansion produced by reserves passing from bank to bank.

[11] If this ratio is unpredictable, the central bank must use a feedback procedure that offsets random changes.

monetary base and short-term Treasury securities. Through open market purchases that increase this total, the central bank increases expenditure by giving the public an excess of liquid assets relative to illiquid assets. Even with a zero short-term interest rate, open market purchases endow the central bank with the power to create money and control expenditure.

When short-term interest rates are zero, interest-rate-targeting procedures are problematic. The zero floor on market interest rates can prevent the central bank from countering expectations of deflation and from responding to a fall in the natural rate. This zero-bound problem is especially acute when expected deflation turns the zero short-term interest rate into a positive real rate. In this situation, predictable control of money and prices requires that the central bank abandon interest-rate-targeting procedures for reserves-targeting procedures. A reserves target continues to allow the central bank to create money to stimulate portfolio rebalancing and expenditure.

2. BOJ OPERATING PROCEDURES

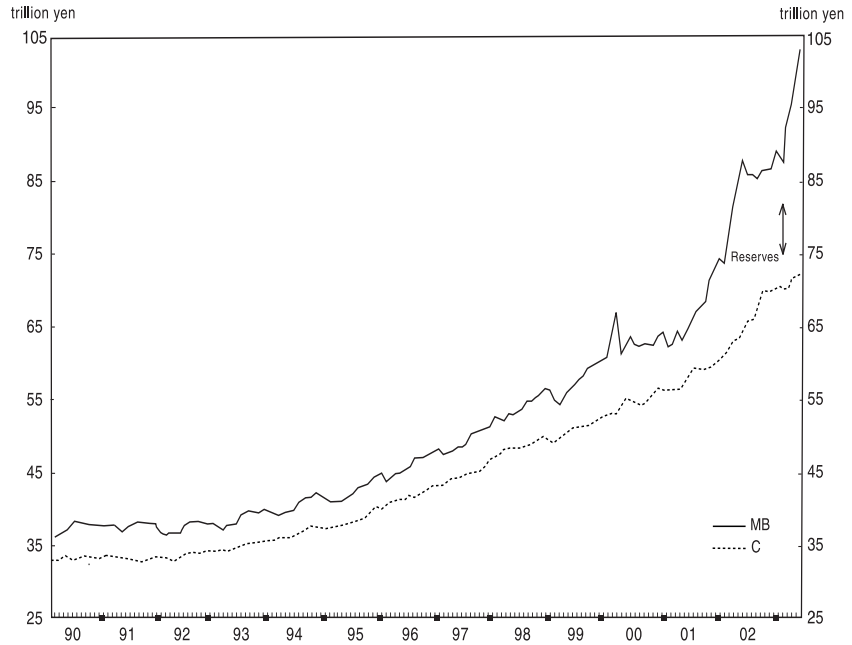
On 19 March 2001, the BoJ began to announce "targets" for reserves balances (current account balances, CABs) held with it by banks. Nevertheless, the BoJ continued to use interest-rate procedures by setting a "target" for reserves equal to estimated reserves demand at a zero overnight call market rate.[12] (The Appendix documents statements in this and the following paragraph.) During the period of the original zero rate policy, February 1999 to August 2000, CABs had averaged 5 trillion yen. With the procedures announced on 19 March 2001, the BoJ adopted this 5 trillion yen figure as a way of reestablishing a zero overnight rate (see Figure 1).

Bank reserves remained demand-determined by the market rather than supply-determined by the BoJ. After 19 March 2001, the BoJ increased its "target" for CABs only in line with increased demand by banks. Demand increased in part because of heightened financial market uncertainty. After 9/11 and after withdrawals from mutual funds following Enron's difficulties, uncertainty increased.[13] On 19 December 2001, the BoJ set a range for CABs of 10 to 15 trillion yen. It used a range because of uncertainty over reserves demand at a zero interest rate. The BoJ commented, "It was a challenge for the Bank of Japan to maintain a high level of current account balances throughout FY2001" (2002, 5). There is no "challenge" if the BoJ determines the amount

[12] The Japanese overnight call money market is comparable to the funds market in the United States. It is, however, open to financial institutions other than banks.

[13] See discussion in *BoJ Quarterly Bulletin* (November 2001 [Minutes, 18 September 2001], 68); *BoJ Quarterly Bulletin* (February 2002 [Minutes, 18–19 December 2001], 94); and Yamaguchi (2002, 36).

Figure 1 Currency and the Monetary Base



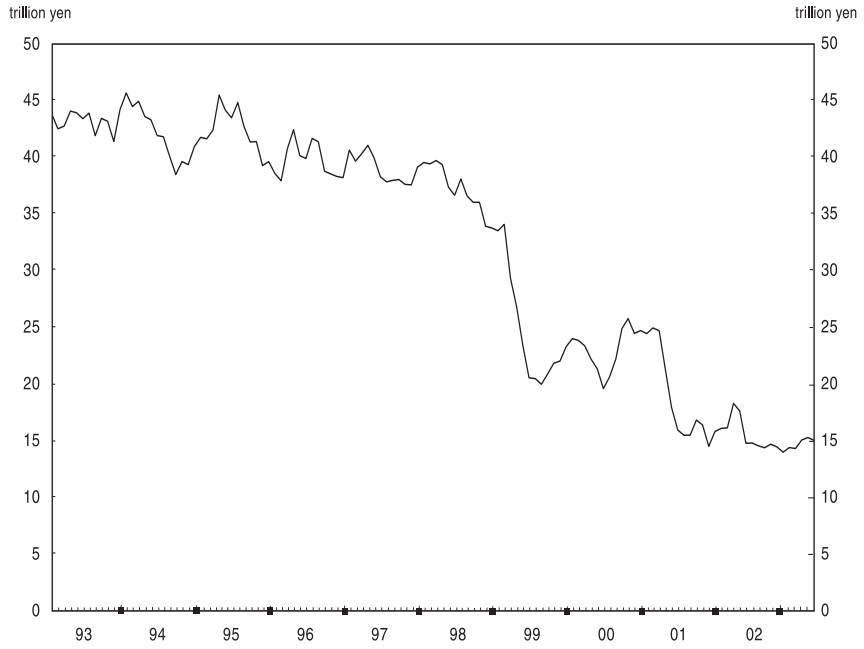
Notes: Monthly observations of currency notes in circulation and the monetary base. Heavy tick marks indicate twelfth month of year. Source: BoJ/Haver Analytics.

of assets to acquire with a reserves strategy instead of limiting itself to the amount of assets necessary to supply the reserves demanded by banks.

The demand for bank reserves has increased because, with a zero short-term interest rate, holding excess reserves becomes an attractive substitute for active reserves management through the use of the call money market. (The call money market provides liquidity by making available overnight loans to meet reserves deficiencies.) On the one hand, excess reserves offer a return equal to the (negative of the) deflation rate. On the other hand, they allow banks to save on the personnel cost of reserves management associated with the use of the call money market (Nakahara 2001, 13). Most important, the use of the call money market began to entail credit risk starting in fall 2001.[14] From 1992 through 1997, total loans in the call money market averaged in

[14] For example, an article in the *Nikkei Weekly* (17 March 2003) states: "The market remains unable to dispel concerns about the risk of a chain of failures of life insurers and banks that could be triggered by the slumping stock market."

Figure 2 Total Average Outstanding Loans in the Call Money Market



Notes: Monthly observations of average outstanding loans in the call money market. Heavy tick marks indicate twelfth month of year. Source: BoJ.

excess of 40 trillion yen. This figure fell dramatically when the BoJ went to its zero interest rate policy in February 1999 (see Figure 2). By mid-2002, it was only 15 trillion yen.

The high level of bank reserves associated with near-zero interest rates creates the misimpression that the BoJ has tried but failed to make banks expand their asset portfolios and deposits. For example, newspapers state that the Japanese banking system is "awash in liquidity." However, individual financial institutions have only substituted the liquidity of excess reserves for the liquidity formerly offered by the overnight call money market (Kodama 2002). The resulting belief that banks have simply impounded reserves supplied by the BoJ generates the mistaken assumption that altering the composition of its asset portfolio, say, by purchasing long-term government bonds (JGBs) is one of the few policy options open to the BoJ. For example, Otsuma (2003a) states, "Buying bonds from commercial banks is one of the few policy tools left to the central bank."

Under the current BoJ demand-driven procedures for reserves provision, the effect of purchases of JGBs and other illiquid assets such as equities are sterilized and thus do not augment total bank reserves. Regardless of whether the interest rate target is positive or zero, bank reserves continue to be demand-determined. A purchase of a JGB, therefore, requires the sale of a short-term security and leaves bank reserves unchanged.[15] To date, the BoJ has limited reserves creation to the amounts demanded by banks. It has not created the additional liquidity that would force an expansion of money.

3. THERE IS NO LIQUIDITY TRAP

With a liquidity trap, the public simply hoards the money the BoJ creates rather than attempting to run down additions with increased expenditure. However, limitless accumulation of money by the public is not a real world phenomenon. The public will not forever accumulate money, which it can use to satisfy real needs.[16]

In Japan there is no evidence for a liquidity trap.[17] Figure 3 shows actual percentage changes in real money (M2+CDs) and the fitted values from the regression in Table 1. Recent real money growth is somewhat stronger than predicted by the regression, but there is no mushrooming demand indicative of a liquidity trap.[18]

In contrast to M2+CDs, M1 growth has risen sharply. M1 growth, which had been around 5 percent, rose in 1995 and then fluctuated around 10 percent until early 2002. At that time, rapid growth in demand deposits (see Figure 4) raised M1 growth to 30 percent. Table 2 shows an M1 demand regression comparable in form to the regression in Table 1. The regression predicts

---

[15] The BoJ open market desk supplies reserves with two sorts of operations: outright purchases and offers (tenders) to sell a specified amount of reserves at the interest rate target, say, zero. If the desk purchases outright a JGB without reducing the offered amount in the latter tender operations, the offer will be undersubscribed. That is, the bid-to-cover ratio will be less than one. Total reserves will remain at the amount demanded at the zero rate of interest. "[W]hether the Bank provides reserves from its right pocket (short-term operations) or from its left pocket (long-term government bond operations), the amount individual financial institutions intend to hold will not change" (Shirakawa 2002, 13).

Purchases of JGBs are comparable to Operation Twist, begun by the Fed in 1961. The Fed purchased long-term government bonds while selling Treasury bills. The idea was to lower bond yields without having to lower short-term interest rates and exacerbate the balance-of-payments problems. Similarly, under current procedures, purchases of JGBs are like central bank purchases of foreign currency in a sterilized foreign exchange intervention. The idea of such intervention is to limit appreciation of the country's currency by altering investors' portfolios to increase the share of domestically denominated assets. However, with no change in the central bank's interest rate target, the monetary base and money remain unchanged (see Broaddus and Goodfriend 1995).

[16] Bernanke (2000, 158) argues: "The monetary authorities can issue as much money as they like. Hence, if the price level were truly independent of money issuance, then the monetary authorities could use the money they create to acquire indefinite quantities of goods and assets. This is manifestly impossible."

[17] Wolman (1997) finds no evidence of a liquidity trap in the U.S. Depression.

[18] The estimates go through 2001, which is the last year for which SNA wealth data are available. (SNA is the System of National Accounts, which in the United States is referred to as the National Income and Product Accounts, or NIPA.)

Table 1 Real Money (M2+CDs) Demand Regression, 1959–2001

$$\Delta \ln M_t = \underset{(2.8)}{.21}\, \Delta \ln M_{t-1} + \underset{(4.2)}{.58}\, \Delta \ln GDP_t - \underset{(5.4)}{.08}\, \Delta \ln(R_t - RM_t)$$

$$+ \underset{(4.0)}{.31}\, \Delta \ln W_t - \underset{(3.8)}{.47}\, E_{t-1} + \hat{\mu}$$

$$CRSQ = .86 \qquad SEE = 1.8 \qquad DW = 1.8 \qquad DF = 38$$

Notes: The regression is in error-correction form. Observations are annual averages, except for wealth, which is a year-end observation. M is M2+CDs divided by the GDP price deflator; R is a rival interest rate paid on nonmonetary assets; RM is a weighted average of the own rates of return paid on the components of M2; W is wealth. E is the estimated residual from a money demand regression in level form using as independent variables GDP, $(R - RM)$, and W; ln is the natural logarithm; Δ is the first-difference operator. $CRSQ$ is the corrected R squared; SEE is the standard error of estimate; DW is the Durbin-Watson; and DF is degrees of freedom. Absolute value of t-statistics is in parentheses.

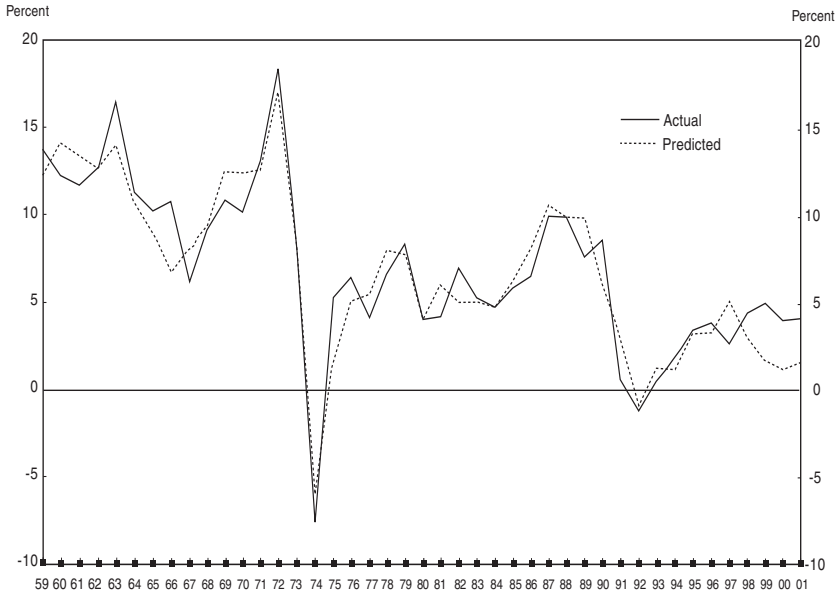
The dates for the regression are determined by the availability of data on the components of M2. Wealth data are available with a one-year lag. The Cabinet Office puts together wealth and GDP data.

From 1957 through 1965, the rival rate (R) is the interest rate on discounts of government securities by banks with the BoJ (boj.or.jp/en/siryo/siryo_f.htm). Thereafter, it is the series used by Toshitaka Sekine (1998) and kindly updated by him. R is the highest interest rate from among the following instruments: three-month (Gensaki) RPs; five-year money trusts; five-year loan trusts; five-year bank debentures (subscription and secondary market); five-year postal savings; and three-year postal savings. The own rate on money (RM) is a weighted average of the own rates on the components of money (demand deposits, time deposits, savings deposits, and CDs).

changes in real M1 until 2001. If there has been an M1 liquidity trap, it is in 2002.

However, the rapid growth in demand deposits in early 2002 reflects a switch from time deposits made in response to a change in government deposit insurance guarantees. In 1996, the government abandoned insurance guarantees limited to 10 million yen on individual deposits for complete coverage. It did so to protect small banks threatened with withdrawals after failure of housing loan corporations (*New York Times,* 23 January 2002). In April 2002, it reimposed the earlier limits by insuring time deposits only up to 10 million yen, while demand deposits remained fully covered. With no interest paid on either type of deposit, depositors could receive unlimited free insurance by switching from time deposits to demand deposits.

Figure 3 Actual and Predicted Real Money Growth



Notes: Predicted values are the within-sample simulated values from the regression shown in Table 1.

4. CAN THE QUANTITY THEORY EXPLAIN JAPANESE DEFLATION?

After 1990Q3, money growth fell by 12 percentage points (see Figure 5).[19] This decline contrasts with the secular rise in real purchasing power demanded. Figure 6 expresses purchasing power as the fraction of nominal output the public holds in money balances. At present, the Japanese hold an amount of money sufficient to fund $1\frac{1}{3}$ times (133 percent of) a year's expenditure on national output. As shown by the trend line, on average, real purchasing power grows by 1.9 percent a year.

What variable has reconciled this fall in the growth of nominal money with the persistent secular rise in the public's demand for purchasing power? A fall in yen output growth in line with money growth maintained desired purchasing power. Over the period 1980Q1 through 1987Q1, money growth averaged

[19] References to money are to M2+CDs. (CDs comprised 2.8 percent of M2 in the ten years after 1992.) I concentrate on M2+CDs rather than M1 because of the more stable demand function for the former than the latter. Note the smaller standard error of estimate of the real M2+CDs demand regression in Table 1 than in the real M1 demand regression in Table 2.

Table 2 Real M1 Demand Regression, 1959–2000

$$\Delta \ln M_t = \underset{(4.5)}{.37}\, \Delta \ln M_{t-1} + \underset{(1.7)}{.29}\, \Delta \ln GDP_t - \underset{(4.9)}{.08}\, \Delta \ln(R_t - RM_t)$$

$$+ \underset{(3.0)}{.31}\, \Delta \ln W_t - \underset{(2.3)}{.14}\, E_{t-1} + \hat{\mu}$$

$$CRSQ = .73 \qquad SEE = 2.7 \qquad DW = 2.1 \qquad DF = 37$$

Notes: See Table 1. M is real M1 (M1 divided by the GDP price deflator). R is a rival interest rate paid on the non-M1 components of M2. It is a weighted average of the own rates paid on time deposits and CDs. RM is a weighted average of the own rates of return paid on the components of M1. The data are from Toshitaka Sekine (1998) and have been kindly updated by him.

8.3 percent, while nominal GDP growth averaged 5.9 percent. (This period serves as a natural benchmark because the money growth during it produced the near price stability of the mid-1980s.) Over the period 1990Q3 to 2002Q3, money growth averaged 2.8 percent, while nominal GDP growth averaged 0.9 percent. Between the two periods, money growth fell 5.5 percentage points and nominal output growth fell almost the same amount, 5 percentage points.

Nominal output is real output measured in yen. Nominal output can change because either real output or the price level changes. In the 1990s, initially, real output growth fell and then prices. Disinflation turned into deflation.

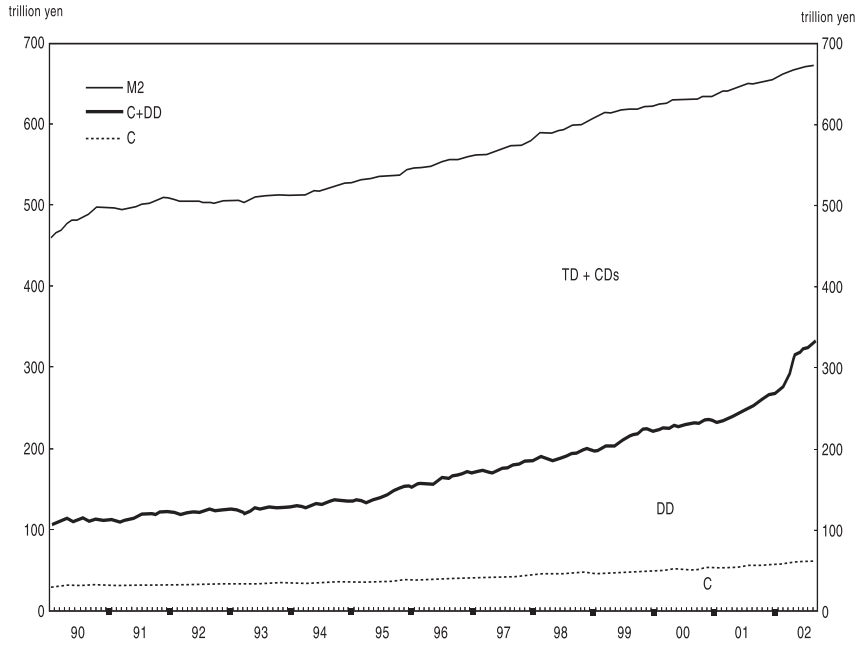
5. A QUANTITATIVE STRATEGY TO STABILIZE THE PRICE LEVEL

Even with a zero interest rate, a central bank can still create money.[20] A strategy based on the price level as the target and bank reserves as the instrument would stimulate yen expenditure by inducing portfolio rebalancing.[21] In this section

[20] Economists arguing that the BoJ should undertake aggressive open market purchases to end deflation include Bernanke (2000), Friedman (1997), Goodfriend (1997), Krugman (1998), and Meltzer (1998). McCallum (1992) argues that the BoJ should use the monetary base as an instrument to control growth of nominal output.

[21] A monetary policy strategy involving reserves-aggregate targeting would require Japan to move to a system of contemporaneous reserves accounting when excess reserves have fallen to normal minimal amounts. At present, Japan has partially lagged reserves accounting. Banks calculate their required reserves based on the daily average of their deposits over a month. The reserves settlement period runs from the 16th of a particular month to the 15th of the following month. Instead of adopting contemporaneous reserves accounting, the BoJ could set required reserves ratios

Figure 4 Currency, Demand Deposits, Time Deposits Plus CDs, and M2



Notes: Monthly observations. C = currency; DD + C = demand deposits + currency; M2 = DD + C + (time deposits + CDs). Seasonally adjusted using RATS esmooth command. Heavy tick marks indicate twelfth month of year. Source: BoJ/Haver Analytics.

I explain this power for two such strategies under the assumption of zero short-term interest rates. The first is a pure transfer of money. The second involves an open market purchase of an asset.[22]

at zero. The need to hold reserves to clear payments would then determine reserves demand. A 1957 law establishing reserves requirements makes this latter option unlikely.

[22] As a third alternative, the BoJ could "target" the term structure of interest rates. Bending the term structure down would be stimulative, and conversely. The BoJ cannot actually peg a long-term interest rate because it cannot credibly commit to maintaining the implied pattern of future short-term interest rates. For example, on 4 March 2003, the implied one-year-forward rate four years into the future was about 0.5 percent (BoJ 2003). Targeting a reduced interest rate on a four-year bond would in principle require committing to making short-term interest rates four years into the future less than 0.5 percent.

Aiming for a less steeply sloped yield curve would force monetary base creation through the tension created between the implied pattern of forward yields and the pattern expected by the public. However, the amount of base money created would be highly unpredictable. To avoid this "shotgun" approach, the BoJ could simply decide on the amount of base money to create based on the extent of the price level target miss.

Outright Money Transfers

With the first strategy, the central bank increases base money by crediting the deposit account the Treasury holds with it.[23] The Treasury delivers such increases to individuals as outright transfers (in a way unrelated to their existing money holdings). After the increase in money, the public still holds no additional liquidity because, at a zero interest rate, the marginal value of the liquidity that economizes on transactions is zero. However, the public now holds purchasing power in excess of what it desires. Because wants are unlimited and the additional money serves no useful purpose, individuals will spend it either on consumption or on acquiring nonmonetary (illiquid) assets. Only a rise in the price level can restore equilibrium.[24]

A real balance effect stimulates expenditure. Increases in base money increase the public's wealth. Increases in this monetary wealth are savings. Because the public saves more in monetary form, it saves less in a nonmonetary form. Consequently, its expenditure rises (Friedman 1976, 320).

Open Market Operations to Increase Money

In practice, central banks create base money through open market purchases in which the public gives up a financial asset in return for a bank deposit. With a zero short-term interest rate, the purchase by the central bank of a Treasury bill does not increase liquidity because Treasury bills are perfectly substitutable for money. No incentive for portfolio rebalancing arises. If the central bank purchases an asset imperfectly substitutable for money, it can force portfolio rebalancing just as in the outright transfer example.[25]

Two related issues arise with zero short-term interest rates. Which assets are imperfect substitutes for money and what magnitude of open market purchases must the central bank undertake to induce portfolio rebalancing? The magnitude of the open market purchases required to produce portfolio

---

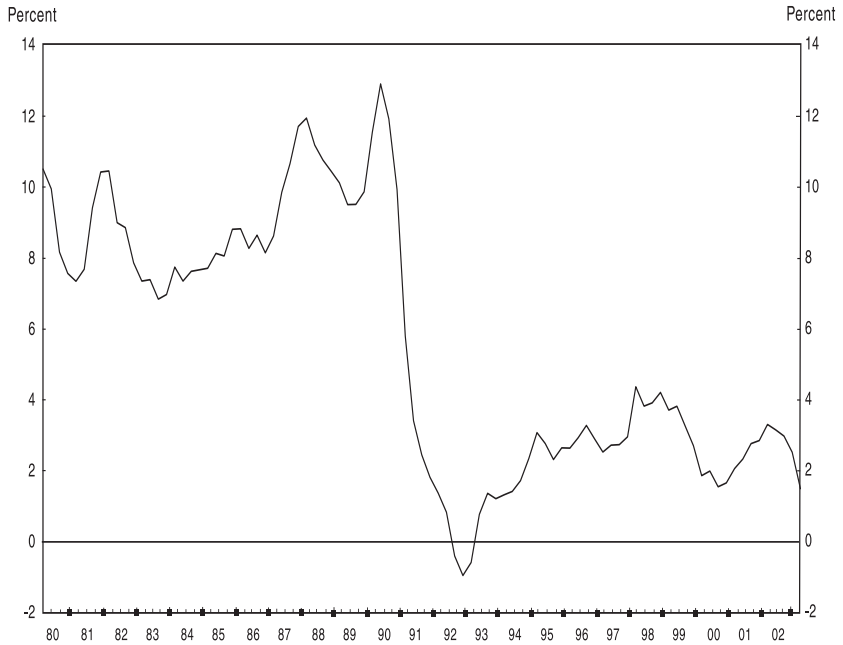
[23] A variation would be for the central bank to purchase JGBs through open market operations in amounts sufficient to provide for increases in currency. Beyond those amounts, it could credit the Treasury's demand deposit to increase the monetary base.

[24] Imagine a government budget constraint in real terms relating the deficit to the issue of government bonds and seigniorage (the increase in nominal money divided by the initial price level). As long as government commits to maintaining a given fiscal policy, the only variable left to adjust to the increase in nominal money is the price level.

[25] Milton Friedman wrote the author (15 April 2003):

> In the preceding case [Outright Money Transfers], the transfer of money raises total nominal wealth. It is windfall income and recipients are inclined to spend at least part of it. In addition, they have been made temporarily to hold a distribution of assets that is not their equilibrium distribution. In the second case in which the central bank operates by purchasing assets by open market operations, the effect is limited to the rebalancing of an improperly structured portfolio.

Figure 5 Money Growth



Notes: Quarterly observations of four-quarter percentage changes of money (M2+CDs). Heavy tick marks indicate fourth quarter of year. Source: BoJ/Haver Analytics.

rebalancing depends upon whether assets like JGBs, corporate bonds, and equities are good substitutes for money. Although in theory these assets could be perfect substitutes, the possibility is highly implausible.[26]

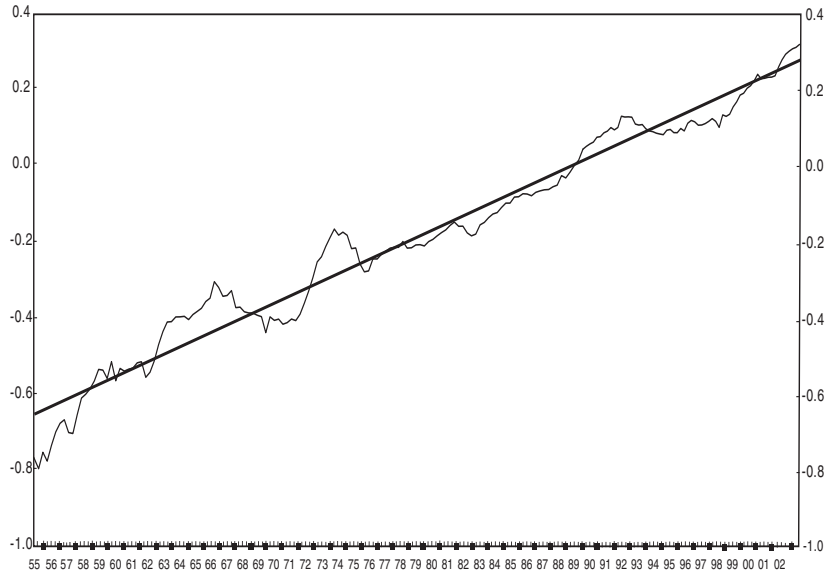
What can one say about the likely magnitude of the asset acquisition required of the BoJ to stimulate the public's yen expenditure? Goodfriend (2000, 2001) argues that to spur expenditure with a zero short-term interest rate, the central bank needs to expand "broad money." He distinguishes between the liquidity services offered by narrow money and broad money. With a zero

[26] At a zero interest rate, money is a perfect substitute for a short-term bill. A short-term bill *could* be a perfect substitute for a JGB. By the expectations hypothesis, a long-term interest rate is just an average of short-term rates. An individual could be indifferent between holding a succession of three-month bills and a JGB. The higher interest rate on the JGB could simply reflect the market's expectation that short-term interest rates will rise in the future.

Furthermore, JGBs can be a perfect substitute for equity. Imagine that the government issues debt and uses the proceeds to purchase equities. If individuals understand that the government is simply holding the equity for them, their behavior is unchanged. It follows that if money is a perfect substitute for JGBs, it is a perfect substitute for equity.

This complicated chain of reasoning pushes the logical limits of what one can assume about investor preferences. Brunner and Meltzer (1968) and Meltzer (1999) question the idea that all financial assets become perfect substitutes at a zero interest rate.

Figure 6 The Demand for Real Purchasing Power



Notes: Quarterly observations of the natural logarithm of M2/GDP with trend. The solid line is the trend line derived from the fitted regression $\ln(M2/GDP)*400 = -263 + 1.9T + \hat{\mu}$. T is a time trend. Heavy tick marks indicate fourth quarter of year. GDP is SNA68 through 1979, SNA93 thereafter. Source: Cabinet Office/Haver Analytics.

interest rate, the public is sated with the transactions services offered by narrow money, but not with the liquidity services offered by broad money that facilitate financial intermediation in a world of agency problems and asymmetric information between borrowers and lenders. For example, because the assets included in broad money are useful as collateral, they lower the cost of credit to a borrower by lowering the finance premium required for external finance. A central bank can increase broad money and spur expenditure by increasing broad money through open market purchases.

For Japan, M2+CDs constitutes a measure of broad money. In 2002Q4, it comprised 134 percent of GDP.[27] A 6 percent rate of increase in M2+CDs,

[27] For the United States, M2 was 55 percent of GDP in 2003Q4. The Japanese save more in the form of money than Americans do. At the end of March 2002, Japanese households held 54.1 percent of their assets in the form of currency and bank deposits. The corresponding figure

a rate consistent with price stability, would not require vast increases in the BoJ's asset portfolio.[28]

The explicitness of a reserves-aggregate strategy for controlling prices would shape expectations of inflation in a way that reinforces the effects of money creation. Such a strategy entails not only a procedure for altering reserves in response to misses in the target for prices, but also an explicit numerical target for the price level.[29] If the price level falls below the targeted price level and policy is credible, the public will expect the price level to rise to eliminate the target shortfall. Expected inflation will raise market rates of interest and reduce money demand. Money demand falls at the same time that the central bank increases money. The public will rebalance its portfolio to eliminate the resulting excess of actual over desired money.

Monetary Indicators

Friedman (1960) argued that lags between central bank actions and changes in the price level make targeting the price level directly destabilizing. He suggested targeting steady money growth to avoid the problem of "long and variable lags."[30] As a supplement to a price level target, the BoJ could use

---

for U.S. households was 11.6 percent (BoJ, *Flow of Funds Accounts;* Board of Governors, *Flow of Funds Accounts of the United States*).

[28] The low value of the reserves-money multiplier limits the required magnitude of the increase in base money. In February 2003, the monetary base equaled 14 percent of M2 and CABs 3 percent. To take an illustrative example, with a constant monetary base/money ratio, a 20 percent increase in money (M2+CDs) would require a 20 percent increase in the monetary base. (In the following, T¥ indicates trillion yen. All figures are for February 2003.) With a monetary base of 93 T¥ and money of 672 T¥, a 20 percent increase in each would amount to 18.6 T¥ and 134.4 T¥, respectively. This increase in the monetary base amounts to 3.4 percent of 2002Q4 GDP.

The required increase in the monetary base would be less if the public held the proceeds from its asset sales to the central bank only in deposits rather than adding to currency (the ratio of currency to total commercial bank deposits is 12 percent). However, the increase in the base would rise if the public held all of the increase in deposits in demand deposits rather than time deposits and banks held reserves primarily against demand deposits. (The ratio of demand deposits to total bank deposits is 0.45. The ratio of reserves to demand deposits is 0.074.)

[29] An inflation target set at a positive rate is a promise to make the currency lose some of its purchasing power each year. A target for the price level is a promise to maintain the purchasing power of the currency. Price indices are biased measures of inflation because they do not account for change in quality. Shiratsuka (1999) places the bias for Japan at 0.9 percent per year. A target path for the CPI price level consistent with genuine price stability would then rise at about 1 percent a year.

Summers (1991) argues that central banks should maintain a positive inflation rate to avoid the zero-bound problem. However, a credible target for the price level would work better. Wolman (1998, 16) points out that when the price level falls below target, such a target ensures a reduction in real rates through transitory increases in expected inflation.

[30] Goodfriend and King (1997, 273–74) argue that a central bank can stabilize the price level with a reaction function that makes its policy instrument vary directly with the discrepancy between the actual price level and the targeted price level. They argue that with credibility price setters will be "forgiving" of policy mistakes. However, this credibility, along with the staggered price setting assumed by Goodfriend and King, implies that a central bank can run, say, an expansionary monetary policy for a very long time before the price level rises.

money and nominal expenditure growth as indicator variables to aid in setting its reserves-aggregate instrument.

Because of the considerable stability in the public's money demand function, money (M2+CDs) has been a better indicator of the thrust of monetary policy than interest rates.[31] The BoJ set the overnight rate at 0.5 percent in September 1995 and at almost zero in February 1999. Low money growth has been a better predictor of deflation than "low" interest rates.

If money demand did become unstable with a reserves-aggregate strategy, the BoJ could use the yen expenditure of the public as an intermediate target. Price stability requires yen expenditure growth equal to sustainable real growth. The BoJ could set a target for yen expenditure growth equal to its estimate of trend real output growth.[32]

6. DEPRECIATION OF THE EXCHANGE RATE

Twentieth-century experiments with fiat money have validated the central implication of the quantity theory that the central bank determines the behavior of the price level. A corollary is that when the central bank pegs the exchange rate, the domestic price level varies to equilibrate the balance of payments. With the exchange rate fixed and foreign prices given, domestic prices must vary to price domestic goods in a way that achieves balance on the external account. Proposals for ending Japanese deflation through a depreciation of the yen build on this fact.[33]

Consider hypothetical yen depreciation achieved by the abandonment of floating exchange rates. For example, the BoJ could peg the yen-dollar

---

For example, monetary policy was expansionary in the United States after 1964 and in Japan after 1997 for two years before inflation rose. From its experience, the BoJ drew the conclusion that a central bank should look at asset prices rather than at the price level (Hetzel 1999). However, the relationship between asset prices and the price level is nebulous. In both the U.S. and Japanese cases, the central banks would not have allowed inflation if they had looked at money.

[31] For time-series studies demonstrating stability, see Sekine (1998) and Bank of Japan (1997). For cross-sectional data indicating stability of money demand, see Fujiki (2002) and Fujiki et al. (2002).

[32] Economic recovery (elimination of a negative output gap) would require an initial target higher than the long-run trend growth of real output. Also, deflation may have reduced output growth over the last decade. In that case, the 1974–1991 average for productivity growth of 2.7 percent is likely to be a better estimate of trend output growth than the 1992–2002 figure of 1.1 percent. (Productivity is calculated as real GDP per worker.) From 1999 through 2001, the population aged 20 to 59 declined 0.4 percent per year. This secular decline in the labor force lowers trend output growth.

[33] Several economists have advocated the use of an exchange rate target to escape the zero-bound problem. Svensson (2001) advocates an initial yen devaluation to end Japanese deflation followed by price level targeting. This policy would raise the Japanese price level while still tying down the public's expectation of the future price level. McCallum (2000, 2003) argues that the BoJ should use the exchange rate as an instrument to control prices.

The BoJ's current interest-rate procedures result in the sterilization of its purchases of dollars. For example, following instructions from the Ministry of Finance, the BoJ purchased 4 trillion yen ($33 billion) in May and June 2002. The BoJ sterilized the resulting reserves creation. That is, it did not allow bank reserves and the money stock to rise.

exchange rate at 150, a 25 percent depreciation from the end-2002 value of 120. In itself, this action will induce a 25 percent rise in the price of traded goods in Japan. The public's decisions determining foreign trade depend upon the relative price of Japanese goods in terms of foreign goods, that is, the terms of trade. A 25 percent depreciation of the yen requires a 25 percent rise in Japanese prices to reestablish the former terms of trade.[34]

Because the Ministry of Finance possesses legal responsibility for the foreign exchange value of the yen, a policy of yen depreciation to control inflation would endanger BoJ independence. Given the large and increasing amount of government debt, financial markets could become concerned that an end to independence might lead to pressure to monetize government debt regardless of the consequences for inflation. Furthermore, Japan must always deal with the protectionist proclivities of its trading partners. A policy of yen depreciation would poison its relations with other countries.[35]

7. INSTITUTIONAL CONSTRAINTS ON THE BOJ

Under the 1998 law establishing central bank independence, Policy Board members are responsible for the "solvency" of the BoJ. Specifically, the BoJ retains 5 percent of its earnings for capital and pays the remainder to the government. At present, the BoJ's capital amounts to 7.6 percent of its assets. The BoJ is concerned that increasing money sufficiently to stop inflation will require not only a large increase in its asset portfolio, but also an increase in nontraditional risky assets.[36] Governor Toshihiko Fukui said, "The institution

[34] After the depreciation, Japanese goods are 25 percent less expensive to foreigners. The BoJ finances the additional demand for Japanese exports by placing newly created yen in the hands of foreigners in return for dollars. Foreigners exchange those yen for Japanese goods, while the Japanese exporters use their newly acquired yen to purchase Japanese securities or deposit the funds in banks. Either way, the Japanese money stock rises. Japanese producers will not forever surrender real resources for low-yielding financial assets. Instead, they will attempt to reduce their money balances through increased spending. Only a rise in the Japanese price level sufficient to restore the former terms of trade can eliminate this imbalance.

As a byproduct of the depreciation, the trade deficit increases transitorily while the price level rises to restore the equilibrium terms of trade. A "large" rise in the price level does not require a "large" trade deficit. Using a model simulation for Japan, McCallum (2003) shows that a devaluation of the Japanese yen need not entail a large trade deficit. The reason is that the stimulative effect of the devaluation also increases imports. It is wrong to argue that "for depreciation to have any real impact on price levels…the yen would have to fall by a huge amount…because trade accounts for a relatively small proportion of the Japanese economy" (Fidler and Guha, *Financial Times,* 23 November 2001). (In 2002, Japan's exports amounted to about 10 percent of GDP.)

[35] Twice the United States pursued a policy of dollar devaluation. The first time was in March 1933, when it devalued the dollar in terms of gold—a policy termed "beggar thy neighbor." The second time was in August 1971, when President Nixon imposed an import surcharge as a club to force countries to revalue their currencies (devalue the dollar). See Hetzel (1999, 2002). Each instance engendered resentment among U.S. trading partners.

[36] Ideally, from an economic perspective, the BoJ would have to increase the size of its asset portfolio significantly to expand the money stock sufficiently to end deflation. In that way, the

that can take indefinite risks is the government alone. Central banks can't go ahead limitlessly—we should never, ever forget this point."[37]

The BoJ is concerned that a fall in the market value of its assets could erase its capital account. The solution to these institutional concerns is political. The government could promise to simply transfer (deliver without monetization) to the BoJ the amount of government securities required to maintain the value of its capital account. The BoJ could then expand its asset portfolio by acquiring assets whose prices fluctuate.

While these concerns may well be determining for BoJ policymakers, it is still important to put them into an economic context. The terminology of "solvency" can possess legal implications for a central bank, but it is not a meaningful economic concept. The economic issue is how the central bank uses the seigniorage from money creation.[38]

It is important for central banks not to attempt to allocate credit by purchasing private securities, especially of insolvent institutions (Hetzel 1997). For the central bank of a less developed country that cannot restrict borrowing by insolvent banks, the problem is real. The central bank may have to monetize so much debt that it creates inflation. However, Japan is not in this situation.

If the BoJ did decide to expand its asset portfolio by purchasing assets other than short-term government debt, it could start with JGBs. In principle, it is possible that ending deflation would require massive open market purchases, which could at a later date require offsetting sales to prevent inflation. The BoJ might then be in the situation of buying JGBs at a high price and selling them later at a low price.[39] The practical import of this situation is that when

---

BoJ could take a large amount of JGBs off the books of banks. When the BoJ does end deflation, interest rates will rise and bond prices will fall. A panic could result if banks collectively attempt to sell bonds. The fall in bond prices could create uncertainty about the solvency of banks. The more long-term bonds that the BoJ has removed from the books of banks, the stronger the financial system will be. The BoJ would then need a transfer of short-term securities from the government to maintain a positive value of its capital account.

[37] The material in this paragraph is from Otsuma (2003b).

[38] Consider the specific example of a central bank lending to an insolvent bank (with no deposit insurance). If the bank fails, the central bank is left with worthless debt, which it writes off. The central bank has purchased private market debt rather than government debt. As a consequence, more government debt ends up in the hands of the public. The real burden of government debt is correspondingly higher. The reason is that interest payments on government debt to the public affect the size of the government deficit. Interest payments to the central bank do not because they are simply recycled. To the extent that the central bank owns government debt, there is no meaningful national debt burden.

[39] If the BoJ is concerned about commercial bank insolvency, it should purchase JGBs from banks to protect them from a future rise in interest rates. Major banks own more than 50 trillion yen in government bonds (*Nikkei Weekly*, 17 March 2003). If the BoJ were concerned about maintaining the market value of its portfolio following economic recovery, it could diversify by buying mutual fund shares holding a diversified selection of stocks.

it comes time to contract the monetary base, it might not have sufficient assets and might have to issue its own debt.[40]

In evaluating BoJ concerns over capital adequacy, one should recognize that the current policy already leads down the path the BoJ wants to avoid. As discussed earlier, the zero short-term interest rates produced by deflation have increased bank holdings of excess reserves by limiting the scope of the call money market. Since March 2001, the value of CABs has risen by about 25 trillion yen. Furthermore, the BoJ is under political pressure to acquire a variety of risky assets such as stocks, to maintain their market value, and securitized business loans, to make funds available to small business. A policy of monetary expansion to end deflation would hold open the prospect of an ultimate solution to the BoJ's capital problems.

For Japanese society, the issue is far more important than the legal and technical one of capital adequacy and the use of seigniorage revenues. First, in the 1990s, disinflation in Japan likely lowered real growth.[41] The fall in asset prices in Japan reflects the reduction in wealth from lower real growth. Second, as I explain in the next section, the zero short-term interest rates produced by expected deflation impede the proper functioning of the price system.

8. ECONOMIC FRAGILITY

To understand why the Japanese economy is now susceptible to adverse shocks, recall the distinction made earlier (in "How a Central Bank Controls the Money Stock") between the real rate and the natural rate. The real rate is the nominal interest rate adjusted for expected inflation (deflation). The natural rate is the real rate that would occur in the absence of monetary disturbances.

The Japanese economy is in a fragile equilibrium because an adverse real shock would simultaneously raise the real interest rate and lower the natural rate. An adverse shock would raise the real rate by increasing expected

[40] The BoJ possesses legal authority to issue debt.

[41] There is evidence to support the contention that the difficulty of adjusting nominal wages to disinflation has lowered real growth. First, during the disinflation in the early 1990s, labor's share of income rose from 65 to 75 percent. The persistence of that elevated share through 2002, despite rising unemployment, indicates incomplete adjustment of nominal wages to lower prices. Second, real wages are not procyclical. Third, adjustment of bonuses has exercised only a limited impact on real wages. Fourth, nominal wages of full-time and part-time workers have remained practically unchanged since the early 1990s (see Fujiki et al. 2001 and Kodama 2001–02). Corporations have adjusted the overall nominal wage by replacing full-time workers with part-time workers, a practice that likely lowers productivity.

deflation. It would lower the natural rate by making the public more pessimistic about the future (Goodfriend 2002). With nominal short-term interest rates equal to zero, the nominal interest rate cannot fall to bring the real rate into equality with a lower natural rate.

Events in October 2002 have already produced this dilemma. In early October 2002, Heizo Takenaka replaced Financial Services Minister Hakuo Yanagisawa. Takenaka desires to prevent banks from lending to insolvent firms. Pessimism about the economy increased from fears that his policy would increase bankruptcies and consequently exacerbate unemployment. Increased pessimism about the future lowered the natural rate.

At the same time, the real rate rose as a result of heightened fears of deflation. Both the Daiwa Institute of Research and the Deutsche Bank economics-forecasting groups predict changes in prices. As of end 2002, both groups forecast a fall of 1.4 percent for the GDP deflator in 2003.[42] Monetary deceleration accompanied this tension in movements in the natural rate and the real rate. In October 2002, year over year money growth was 3.5 percent. By April 2003, it had fallen to 1.4 percent.

At the current deflation rate, this growth in nominal money allows for only minimal growth in real output. Since 2000, inflation (GDP deflator) has averaged -1 percent. Nominal money growth of 1.4 percent then implies 2.4 percent real money growth. The trend growth in real purchasing power is 1.9 percent (see Figure 6), which leaves less than 1 percent real money growth to accommodate real output growth.

9. CONCLUDING COMMENTS

Inflation and deflation are monetary phenomena. They depend upon the way the central bank creates money. The BoJ can end deflation by raising money growth. To do so, it would need to abandon its current policy of limiting base money creation to the amount demanded by the public. Instead, it should adopt an explicit target for the price level and a policy of monetary base creation to achieve that target.

APPENDIX: DEMAND-DETERMINED RESERVES PROVISION

In February 1999, the BoJ adopted a target for the uncollateralized overnight call rate of interest of near zero. In August 2000, it raised its target to 25 basis points. However, the economic recovery that had prompted that rise

[42] Deutsche Bank Group (2002) and Daiwa Institute of Research (2002).

ended that fall. The BoJ then adopted reserves-targeting language allowing it to return to its former zero rate policy without an explicit reversal.

The 19 March 2001 "Minutes of the Monetary Policy Meeting" (*BoJ Quarterly Bulletin,* May 2001, 82) state:

> [T]he effects previously brought about by the zero interest rate policy could be achieved and at the same time the market mechanism could be maintained to some extent, if the operating target was changed to the outstanding balance of current accounts at the bank and the amount was increased to a level that would reduce the interest rate to virtually zero (the level was estimated to be around 5 trillion yen given the experience of the zero interest rate policy). [Expected inflation would not rise] if the quantitative easing was *limited* to the level necessary to achieve a fall in the overnight call rate to virtually zero. (*italics added*)

At the 13 August 2001 Monetary Policy Meeting, the BoJ raised the CAB target to 6 trillion yen, "the *maximum* amount possible" (*italics added*) (*BoJ Quarterly Bulletin,* November 2001 [Minutes, 13 August 2001], 45). There is no "maximum" amount to a reserves target set by the central bank.

After March 2001, the BoJ increased reserves provision only in line with increases in demand by banks. Masaaki Shirakawa (2002, 9), adviser to the governor, explained the increase in reserves that occurred after March 2001 as reflecting factors affecting bank demand for reserves: "an increase in domestic financial institutions' precautionary demand for liquidity against the background of uncertainty with respect to liquidity conditions" and an increase in demand from foreign banks arising from yen-dollar swap transactions. Yutaka Yamaguchi (2001, 6), BoJ deputy governor, explained:

> [T]he Bank did not simply raise the target [for CABs] regardless of demand. The Bank decided the level of the target... based on a judgment that it was maximum demand for the current account balance at the time. In September [2001], the Bank swiftly responded to the surge in demand for liquidity.... [T]he Bank can increase the current account balance flexibly as long as demand for liquidity increases.... The current account balance can be increased when a certain stress gives incentives for financial institutions to hold a larger amount of liquidity.

Policy Board member Nobuyuki Nakahara (2001, 11–12) argued that "[t]he Bank is simply providing funds to accommodate funds demand." He detailed examples of funds absorption by the open market desk to show that the BoJ does not force unwanted reserves on financial institutions. Board member Shin Nakahara (2002, 3) commented, "[T]he outstanding balance of current accounts at the Bank cannot be increased 'without limit' since it cannot exceed the actual demand for funds by financial institutions."

The BoJ has set its "target" for CABs as a range to allow for reductions in bank demand for reserves:

> These members raised the question of whether, if liquidity demand decreased for some reason, the Bank could continue its provision of funds to maintain the outstanding balance of current accounts at a high level. . . . The staff pointed out that. . . depending on liquidity demand, there was a possibility that the total amount of bids in market operations would often fall short of the amount the Bank offered, i.e., a possibility of undersubscription. . . . The Bank should be capable of dealing with the situation where demand for funds decreased as the demand did not seem to have become stable yet. (*BoJ Quarterly Bulletin*, February 2002 [Minutes, 18–19 December 2001], 101)

The fact of "undersubscription" shows that the BoJ limits reserves creation to the amount demanded by banks. "[In FY2001] undersubscription for fund providing operations was not uncommon" (BoJ 2002, 1). "Many members said that the undersubscription was proof that the Bank was providing liquidity to its utmost" (*BoJ Quarterly Bulletin,* May 2002 [Minutes, 7–8 February 2002], 35). The bid-to-cover ratio measures undersubscription. For example, the BoJ's repurchase operations on 2 May and 9 May 2001 were undersubscribed with bid-to-cover ratios of 0.9 and 0.4, respectively (Chen 2001).

This ratio measures the supply of bills the market offers to the BoJ relative to the bills that the BoJ is willing to buy. (The latter figure, the amount that the BoJ is willing to buy, comes from estimates of purchases necessary to provide just enough reserves to maintain the overnight call rate at zero.) The BoJ purchases the former, the amount the market offers, not the latter amount. The bid-to-cover ratio would be irrelevant if the BoJ simply bought the amount of assets required to achieve a given target for bank reserves. "Undersubscription" can occur only if the BoJ allows market demand to determine reserves provision.

With demand-determined reserves provision, the BoJ limits reserves creation to the amount that banks demand at a zero interest rate. With active reserves provision, the BoJ would supply reserves beyond this amount. Bank reserves demand would then increase to match supply because of an increase in bank deposits.

REFERENCES

Bank of Japan. 1997. "On the Relationship Between Monetary Aggregates and Economic Activities in Japan: A Study Focusing on Long-Term Equilibrium Relationships." *Bank of Japan Quarterly Bulletin* (November): 104–24.

———. 2002. "Market Review: Money Market Operations in FY2001" (July).

———. 2003. "Monthly Report of Recent Economic and Financial Developments" (March).

———. *Bank of Japan Quarterly Bulletin*. "Minutes of the Monetary Policy Meeting." Various issues.

———. *Flow of Funds Accounts*. www.boj.or.jp/en/stat/stat_f.htm.

Bernanke, Ben S. 2000. "Japanese Monetary Policy: A Case of Self-Induced Paralysis?" In *Japan's Financial Crisis and Its Parallels to U.S. Experience,* edited by Ryoichi Mikitani and Adam S. Posen. Washington, D.C.: Institute for International Economics,149–66.

———, and Mark Gertler. 1995. "Inside the Black Box: The Credit Channel of Monetary Policy Transmission." *Journal of Economic Perspectives* 9 (Fall): 27–48.

Board of Governors of the Federal Reserve System. *Flow of Funds Accounts of the United States.* www.federalreserve.gov/releases/z1/current/data.htm.

Broaddus, J. Alfred, Jr., and Marvin Goodfriend. 1995. "Foreign Exchange Operations and the Federal Reserve." Federal Reserve Bank of Richmond *Annual Report*.

Brunner, Karl, and Allan H. Meltzer. 1968. "Liquidity Traps for Money, Bank Credit and Interest Rates." *Journal of Political Economy* 76 (July): 8–24.

Chen, Kathryn. 2001. "Japan Money Market Update: Developments as Interest Rates Approach Zero." Federal Reserve Bank of NewYork Market Source, https://marketsource.ny.frb.org (10 May).

Daiwa Institute of Research. 2002. *Japan's Economic Outlook* 135 (Winter): Table 1.

Deutsche Bank Group. Global Markets Research Japan. 2002. *Japan Economic Quarterly* (November): Table 1.

Fidler, Stephen, and Krishna Guha. 2001. "Weakening the Yen." *Financial Times,* 23 November, 12.

Friedman, Milton. 1956. "The Quantity Theory of Money—A Restatement." In *Studies in the Quantity Theory of Money*, edited by Milton Friedman. Chicago: University of Chicago Press, 3–21.

———. 1960. *A Program for Monetary Stability*. New York: Fordham University Press.

———. 1969a. "The Lag in Effect of Monetary Policy." In *The Optimum Quantity of Money and Other Essays,* edited by Milton

Friedman. Chicago: Aldine Publishing Company, 237–60. Reprint, "The Lag in Effect of Monetary Policy," *Journal of Political Economy* 69 (October 1961): 447–66.

_____. 1969b. "The Optimum Quantity of Money." In *The Optimum Quantity of Money and Other Essays*, edited by Milton Friedman. Chicago: Aldine Publishing Company, 1–50.

_____. 1976. *Price Theory.* Chicago: Aldine Publishing Company.

_____. 1997. "Rx for Japan: Back to the Future." *Wall Street Journal,* 17 December, A22.

Fujiki, Hiroshi. 2002. "Money Demand Near Zero Interest Rate: Evidence from the Regional Data." Institute for Monetary and Economic Studies, Bank of Japan, *Monetary and Economic Studies* (April): 25–41.

_____, Cheng Hsiao, and Yan Shen. 2002. "Is There a Stable Money Demand Function Under the Low Interest Rate Policy? A Panel Data Analysis." Institute for Monetary and Economic Studies, Bank of Japan, *Monetary and Economic Studies* (April): 1–23.

_____, Sachiko Kuroda Nakada, and Toshiaki Tachibanaki. 2001. "Structural Issues in the Japanese Labor Market: An Era of Variety, Equity, and Efficiency or an Era of Bipolarization?" Institute for Monetary and Economic Studies, Bank of Japan, *Monetary and Economic Studies* (February): 177–208.

Goodfriend, Marvin. 1990. "Money, Credit, Banking and Payments System Policy." In *The U.S. Payments System: Efficiency, Risk, and the Role of the Federal Reserve*, edited by David B. Humphrey. Boston: Kluwer Academic Publishers, 247–77.

_____. 1993. "Interest Rate Policy and the Inflation Scare Problem." Federal Reserve Bank of Richmond *Economic Quarterly* 79 (Winter): 1–24.

_____. 1997. "Comments." In *Towards More Effective Monetary Policy*, edited by Iwao Kuroda. Tokyo: Bank of Japan, 289–95.

_____. 2000. "Overcoming the Zero Bound on Interest Rate Policy." *Journal of Money, Credit and Banking* 32 (November, Part 2): 1007–35.

_____. 2001. "Financial Stability, Deflation, and Monetary Policy." Institute for Monetary and Economic Studies, Bank of Japan, *Monetary and Economic Studies* 19 (February): 143–67.

_____. 2002. "Monetary Policy in the New Neoclassical Synthesis: A Primer." *International Finance* 5: 165–91.

_____, and Robert G. King. 1997. "The New Neoclassical Synthesis." In *NBER Macroeconomics Annual,* edited by Ben S. Bernanke and Julio Rotemberg. Cambridge: MIT Press, 231–95.

Hetzel, Robert L. 1997. "The Case for a Monetary Rule in a Constitutional Democracy." Federal Reserve Bank of Richmond *Economic Quarterly* 83 (Spring): 45–65.

_____. 1999. "Japanese Monetary Policy: A Quantity Theory Perspective." Federal Reserve Bank of Richmond *Economic Quarterly* 85 (Winter): 1–25.

_____. 2002. "German Monetary History in the First Half of the Twentieth Century." Federal Reserve Bank of Richmond *Economic Quarterly* 88 (Winter): 1–35.

Kodama, Takashi. 2001–02. "Wage Deflation." Daiwa Institute of Research, *Japan's Economic Outlook* 131 (Winter): 40–44.

_____. 2002. "Rethink on Quantitative Easing." Daiwa Institute of Research (2 September).

Koll, Jesper. 2003. "End Zero Interest Rates in Japan." *Asian Wall Street Journal,* 26 February, A7.

Krugman, Paul. 1998. "Setting Sun Japan: What Went Wrong." web.mit.edu/krugman/www/japan.html (11 June).

McCallum, Bennett T. 1992. "Specification and Analysis of a Monetary Policy Rule for Japan." Institute for Monetary and Economic Studies, Bank of Japan, *Monetary and Economic Studies* 11 (November): 1–45.

_____. 2000. "Theoretical Analysis Regarding a Zero Lower Bound on Nominal Interest Rates." *Journal of Money, Credit and Banking* 32 (November, Part 2): 870–904.

_____. 2001. "Should Monetary Policy Respond Strongly to Output Gaps?" *American Economic Review, Papers and Proceedings* 91 (May): 258–62.

_____. 2003. "Japanese Monetary Policy, 1991–2001." Federal Reserve Bank of Richmond *Economic Quarterly* 89 (Winter): 1–31.

Meltzer, Allan H. 1998. "Time to Print Money." *Financial Times*, 17 July, 14.

_____. 1999. "Liquidity Claptrap." *International Economy* (November/December): 18–23.

Miller, Anthony M. 2003. "Deflation Isn't Japan's Biggest Problem." *Asian Wall Street Journal,* 28 February, A7.

Nakahara, Nobuyuki. 2001. "The Japanese Economy and Monetary Policy in a Deflationary Environment." Speech given at the Capital Markets Research Institute, Tokyo, 11 December.

Nakahara, Shin. 2002. "State of Japan's Economy and Policy Measures." Speech given at the Spanish Embassy, Tokyo, 20 February.

New York Times. 2002. "Jitters in Japan for Savers and Banks," 23 January, W1.

_____. 2003. "Japan Tries a New Tack on Economy: Buying Debt." 9 April, W1.

Nikkei Weekly. 2003. "Stock Plunge Threatens Banks," 17 March, 1.

Otsuma, Mayumi. 2003a. "Fukui May Expand BoJ's Role in Reviving Economy (Update 1)." Bloomberg.com (26 March).

_____. 2003b. "Fukui Deflects Calls to Buy Shares in Banks (Update 1)." Bloomberg.com (1 May).

_____, and Kanako Chiba. 2003. "Bank of Japan Increases Stock Purchases from Banks (Update 7)." Bloomberg.com (25 March).

Sekine, Toshitaka. 1998. "Financial Liberalization, the Wealth Effect, and the Demand for Broad Money in Japan." Institute for Monetary and Economic Studies, Bank of Japan, *Monetary and Economic Studies* 16 (May): 35–55.

Shirakawa, Masaaki. 2002. "One Year Under 'Quantitative Easing.'" IMES Discussion Paper Series 2002-E-3, Institute for Monetary and Economic Studies, Bank of Japan (April).

Shiratsuka, Shigenori. 1999. "Measurement Errors in the Japanese Consumer Price Index." Institute for Monetary and Economic Studies, Bank of Japan, *Monetary and Economic Studies* 17 (December): 69–102.

Summers, Lawrence S. 1991. "How Should Long-Term Monetary Policy Be Determined?" *Journal of Money, Credit and Banking* 23 (August, Part 2): 625–31.

Svensson, Lars E. O. 2001. "The Zero Bound in an Open Economy: A Foolproof Way of Escaping from a Liquidity Trap." Institute for Monetary and Economic Studies, Bank of Japan, *Monetary and Economic Studies* 19 (February): 277–312.

Wolman, Alexander L. 1997. "Zero Inflation and the Friedman Rule: A Welfare Comparison." Federal Reserve Bank of Richmond *Economic Quarterly* 83 (Fall): 1–21.

_____. 1998. "Staggered Price Setting and the Zero Bound on Nominal Interest Rates." Federal Reserve Bank of Richmond *Economic Quarterly* 84 (Fall): 1–24.

Yamaguchi, Yutaka. 2001. Remarks at the JCIF International Finance Seminar. www.boj.or.jp/en/index.htm, "Speeches and Statements" (17 October).

———————. 2002. "Central Banking in Uncharted Territory." *Bank of Japan Quarterly Bulletin* (August): 33–40.

The Euro and Inflation Divergence in Europe

Margarida Duarte

In January 1999, eleven European countries abandoned their respective national currencies and monetary independence to adopt a common currency, the Euro.[1] This event, in which several industrialized countries formed a currency union, stands out in modern monetary history by its uniqueness, and in due time, it will allow for a better understanding of the implications of different monetary arrangements among countries. Already, with four years of data available, we can begin to learn from Europe's natural experiment.

In a flexible exchange rate regime, the equilibrium adjustment in the relative price across countries associated with a given country-specific shock results both from movements in nominal prices and from movements in the relative price of the countries' currencies, i.e., the nominal exchange rate. In a currency union, movements in the nominal exchange rate are, by definition, no longer possible, and equilibrium adjustments in the relative price across countries result only from movements in nominal prices.[2] In addition, countries in a currency union can no longer use monetary policy in response to such a shock. The equilibrium adjustment of nominal prices associated with a given country-specific shock reflects, among other factors, not only the degree of asymmetry of the shock but also the degree of integration of the different regions (namely, the mobility of factors of production or the ability

The author would like to thank Andreas Hornstein, Thomas Humphrey, Roy Webb, and Alexander Wolman for helpful comments. This article does not necessarily represent the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] These countries were Austria, Belgium, Finland, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Portugal, and Spain. Greece adopted the Euro in January 2001. Monetary policy in the Euro area has been conducted by the European Central Bank (ECB) since 1999.

The remaining three members of the European Union (Denmark, Sweden, and the United Kingdom) have, so far, decided to maintain their own currencies and monetary independence.

[2] For example, as will be seen later, in response to faster productivity growth in its traded-goods sector (than in the other sectors), a country will experience a real exchange rate appreciation (an increase in its relative price), which in a currency union translates into higher inflation.

to automatically transfer resources across regions). The more asymmetric the shock or the less integrated the different regions in a currency union, the bigger the equilibrium adjustment associated with a given shock. Hence, inflation differentials can be seen as an indicator of regional asymmetries within a currency union.[3]

In this article I document the behavior of inflation dispersion and inflation differentials in Euro-area countries before and after the introduction of the Euro. This documentation supports the main message of the article: that inflation dispersion and inflation differentials (with respect to German inflation) within the Euro area have *increased* after the adoption of the common currency. Moreover, inflation dispersion in the Euro area has been higher than that observed in the United States. Assessing the sources of inflation divergence in the Euro area after 1999 suggests that countries with higher inflation rates tend to have also had higher GDP growth rates and a lower price level when the Euro was adopted. Finally, the variability of the inflation differential with respect to German inflation has tended to increase for most countries after the Euro was adopted.

This article is organized as follows. In Section 1 I briefly review the process leading to the implementation of the European Monetary Union (EMU). In Section 2 I provide a general discussion about currency unions, and in Section 3 I document the behavior of inflation before and after the Euro was adopted using twelve-month core CPI inflation data from the eleven countries that adopted the common currency in January 1999. In the final section I state my conclusions.

1. A BRIEF REVIEW OF THE ROAD TO THE EMU

The process of European integration started shortly after World War II, stimulated by the idea that a unified Europe would help ensure peace. In 1950 Robert Schuman, France's foreign minister, proposed that the coal and steel industries of France and Germany (then West Germany) be coordinated under a single supranational authority. This initiative lead to the European Steel and Coal Community, formed in 1952 together with Belgium, Italy, Luxembourg, and the Netherlands. Building on the success of this organization, the European Economic Community and the European Atomic Energy Community were established in 1957 by the Treaty of Rome. These three organizations were later consolidated in 1967 to form the European Community (EC), known as European Union (EU) since the ratification of the Maastricht Treaty in 1992.

[3] This discussion is closely related to that of optimal currency areas. The theory of optimal currency areas dates back to Mundell (1961), but it gained renewed interest in the last decade with the European project for a currency union. This theory stresses the relative importance of internal factor mobility and external factor immobility in defining the appropriate domain for a currency area.

As the Bretton Woods system became less stable during the 1960s, the European Council decided in December 1969 to pursue the goal of establishing an economic and monetary union in Europe by 1980.[4] A three-phase plan designed by Pierre Werner (then prime minister of Luxembourg) to achieve economic and monetary union within ten years was approved in March 1971, and the first stage, involving the narrowing of currency fluctuation margins, was launched. However, the instability in foreign exchange markets in 1971 and the subsequent collapse of the Bretton Woods system effectively brought the EMU project to a stop until the end of the decade.

In a new effort to establish an area of monetary stability, the European Monetary System (EMS) was created in March 1979.[5] The EMS allowed (initially) for currency fluctuations in a $\pm$ 2.25 percent range around fixed bilateral rates, and it effectively reduced exchange rate volatility among the participating currencies. It wasn't until 1988, however, that a new effort to establish a monetary union was made when the Hanover European Council commissioned a report to Jacques Delors (then president of the European Commission) on the implementation of a monetary union. The resulting Delors Report laid out a three-stage plan for the implementation of a monetary union, culminating with the creation of a single currency. The first stage of this process began in July 1990 and was marked by the dismantling of internal barriers to the free movement of capital.

In February 1992 the European Council signed the Maastricht Treaty, formally establishing the blueprint for economic and monetary integration in Europe. It defined the precise time line for the three stages leading to monetary union and set out the convergence criteria that member states had to pass in order to be eligible to adopt the common currency (the EMU's final stage).

The first stage of the EMU project, already in place, ended in December 1993. The second stage then began with the establishment of the European Monetary Institute (which would later become the European Central Bank—ECB). Its role was to strengthen the coordination of monetary policies among member states and to make the preparations required for a single monetary policy and currency.

The Maastricht Treaty laid out five convergence criteria that member states had to meet in order to enter into the EMU's final stage. These criteria were (1) public budget deficit below 3 percent of GDP; (2) public debt less than 60 percent of GDP; (3) inflation rate within 1.5 percent of the three EU countries with the lowest rates; (4) long-term interest rates within 2 percent of the three

[4] The European Council is composed primarily of the president of the European Commission (the executive body of the EU) and the heads of government of the member states and their foreign ministers.

[5] The participating countries in the EMS were the six countries that formed the EC since its inception, plus Denmark and Ireland (which joined the EC in 1973). The United Kingdom also joined the EC at this date but opted not to participate in the EMS.

lowest interest rates in the EU; and (5) no nominal exchange rate movements outside the EMS's margins for two years. These convergence criteria, which imposed strict fiscal rules and required inflation and nominal interest rates to converge across Europe, conditioned the conduct of both monetary and fiscal policy in the EU countries before the actual adoption of the common currency.

In the spring of 1998 the European Council announced the eleven countries that would enter the EMU's third stage as well as the irrevocable conversion rates between the Euro and each participating currency.[6] The third stage started in January 1999 with the introduction of the Euro as a medium of account. Euro banknotes and coins were put in circulation in 2002.

With the start of the EMU's third stage, member countries abandoned their monetary independence, and monetary policy came under the control of the ECB. Its goal is to maintain medium-term price stability in the Euro area, defined as a year-on-year increase in the harmonized index of consumer prices (HICP) below 2 percent.[7] With the start of the third stage, member countries also committed to the fiscal rules set by the Stability and Growth Pact. This pact establishes a limit of 3 percent of GDP for budget deficits, and it commits member countries to aim in the medium term for budgets that are close to balance or in surplus.[8]

Several countries may join the EMU in the next few years. In one group are the three EU member states still pending political approval in their countries to join the EMU: Denmark, Sweden, and the United Kingdom. Another group includes the countries that are candidates to join the EU in 2004 and are required to meet the Maastricht convergence criteria. These countries are the Czech Republic, Cyprus, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, and Slovenia.

2. CURRENCY UNIONS AND INFLATION DIFFERENTIALS

Monetary and Fiscal Policies in a Currency Union

In a currency union, different countries or regions share a common currency. The issuance of the common currency and the conduct of monetary policy is the

[6] Of the remaining four EU member countries not entering the Euro zone in 1999, Denmark, Sweden, and the United Kingdom chose not to participate, while Greece was viewed, at this stage, as not having fulfilled the necessary conditions for the adoption of the Euro.

[7] The HICP is the weighted arithmetic average of the consumer price indices for the Euro-area countries. The weight of each country is its share of private domestic consumption expenditure in the Euro area.

See Svensson (2002) for a critical evaluation of European monetary-policy strategy.

[8] The Stability and Growth Pact also defines the exceptional conditions under which breaching the 3 percent budget deficit limit can be accepted and establishes how and when fines can be levied against countries that display excessive deficits.

responsibility of a central monetary authority.[9] This institutional arrangement characterizes, for example, the states that form the United States and the countries that form the EMU; the authorities responsible for monetary policy are the Federal Reserve System and the ECB, respectively.

The central bank of a currency union holds assets which may include interest-bearing instruments issued by the governments of the different regions (or countries) or by the federal government, and its liabilities consist of the monetary base for the whole area. The monetary authority adjusts the money supply through the purchase and sale of interest-bearing assets. Note that since the member countries share a common currency, interest-rate parity implies that the nominal interest rate (on assets with similar characteristics) is the same across countries in a currency union. The joint central bank earns seigniorage revenue from issuing the common currency, and this revenue can be freely allocated across countries.[10]

In contrast to monetary policy (which is decided at the central level), fiscal policy is under the control of a member state or country in a currency union. That is, member states or countries maintain a fiscal authority, responsible for the conduct of fiscal policy in their region. This fact does not preclude the existence of a central fiscal authority as well. This is the case, for example, in the United States, where fiscal policy at the federal level involves a large amount of resources. In Europe, the resources involved in fiscal policy at the central level are very small.

Costs and Benefits of a Currency Union

By adopting a common currency, countries eliminate exchange rate risk and the costs of currency conversion.[11] Under floating exchange rate regimes, nominal exchange rates typically exhibit very high variability, with standard deviations in the order of 7 or 8 percent for quarterly data.[12] Obstfeld and Rogoff (2002) compute the welfare cost of exchange rate variability in an explicitly stochastic version of the "new open-economy macroeconomics" framework.[13] In this context, they consider a monetary regime change that eliminates exchange rate variability (by pegging the exchange rate) while maintaining the variance of

[9] I am restricting attention to arrangements in which a group of countries (or regions) shares a common currency and monetary policy is decided by a joint central bank. I am, therefore, abstracting from arrangements in which a country (typically small) adopts the currency of a large anchor country.

[10] Sibert (1994) considers the problem of allocating seigniorage in a currency union.

[11] The European Commission estimated that costs of currency conversion in the European Union amount to 0.4 percent of the area's GDP.

[12] See, for example, the data presented in Chari, Kehoe, and McGrattan (2002).

[13] The "new open-economy macroeconomics" framework was set forth by Obstfeld and Rogoff (1995), and it represents an important workhorse model in international economics. This model introduces nominal rigidities into a two-country general equilibrium model.

world monetary growth constant. For their parameterization, which assumes a low degree of risk aversion, the cost of exchange rate variability is about 1 percent of GDP. This calculation suggests that the welfare losses due to exchange rate movements generated by monetary shocks alone could be large. Furthermore, it reflects only the benefits of eliminating exchange rate risk, that is, of fixing the exchange rate. Adopting a common currency, however, is understood to have other implications, such as deeper market integration, which may entail additional benefits that are beyond the scope of the model in Obstfeld and Rogoff (2002).

An increase in trade volume is typically stressed as an important implication of reduced costs of currency conversion and the absence of nominal exchange rate risk. Several recent empirical studies have investigated the relationship between currency unions and trade. These studies offer a wide range of estimates of the effect of the currency union on trade and suggest that belonging to a currency union/board may lead to an increase in trade with other members by as much as a factor of three. Among these studies, Rose and van Wincoop (2001) estimate that the Euro may lead to an increase in trade in the Euro area of about 50 percent.[14]

By joining a currency area, however, a country forgoes the ability to use monetary policy to respond to region-specific macroeconomic disturbances. The inability to use monetary policy in response to asymmetric shocks can be an important cost of joining a currency union, particularly if asymmetric shocks represent an important source of output fluctuations and if adjustment mechanisms across regions to these shocks are absent. One such margin of adjustment across regions is provided by factor mobility, which allows factors of production to be easily reallocated in response to regional shocks. Another margin of adjustment to idiosyncratic shocks across regions is the automatic stabilization provided by sizeable transfer programs administered at the union level. The federal income tax and unemployment insurance, which automatically transfer resources from booming regions to those in recession, are examples of such programs administered at the federal level in the United States. Europe differs considerably from the United States along these two dimensions: despite the elimination of barriers to the movement of factors, labor mobility within Europe is still lower than in the United States, and unlike the United States, Europe lacks a sizeable system of transfers among states. Finally, countries in a currency union may also incur strategic and political costs in determining the allocation of seigniorage and the conduct of monetary policy.

[14] See Rose (2002) for a review of this literature and for a complete list of references. This paper, in particular, uses meta-analysis for evaluating and combining the disparate estimates from different studies. He finds that the combined estimate implies that a currency union approximately doubles trade.

Price Level Divergence in a Currency Union

Countries in a monetary union share the same currency but need not have the same price level: different regions within the union may have different price levels and experience different inflation rates.[15] The United States, a long-established currency union, provides a benchmark for the magnitude of price differentials in a currency union. Cecchetti, Mark, and Sonora (2002) use consumer price data for nineteen U.S. cities from 1918 to 1995 and find that price level divergences across U.S. cities are large and persistent: annual inflation rates measured over ten-year periods can differ by as much as 1.55 percentage points. Parsley and Wei (1996) use commodity level price data for forty-eight U.S. cities from 1975 to 1992 and find persistent deviations from the law of one price for both traded and nontraded goods.[16] They also find that convergence rates for traded categories are higher than those of nontraded goods or those found in cross-country data.

Deviations in the price level across regions within a currency union can arise from two sources. The first source is deviations from the law of one price for traded goods across regions. The second source is deviations in the relative price of nontraded goods across regions.

Let us consider a currency union with two regions, A and B, and assume that the price index in each region is given by a geometric weighted average of traded- and nontraded-goods prices:

$$p_{i,t} = \alpha p_{i,t}^{N} + (1-\alpha)\, p_{i,t}^{T},\, i = A,\, B,$$

where $p_{i,t}$ is the log of the price index, $p_{i,t}^{T}$ $\left(p_{i,t}^{N}\right)$ is the log of the traded- (nontraded-) goods price index, and α is the share of nontraded goods in the price index.[17] Clearly, asymmetric shocks within a currency union, with distinct effects on the price index of traded or nontraded goods (p^T or p^N) across regions, will generate a differential in the price level across countries and an inflation differential.[18]

One such asymmetric shock is the following. If a country experiences faster productivity growth in the sectors producing traded goods (relative to

[15] Much of the existing literature on monetary unions associates a common currency with a common price level. See, for example, Canzoneri and Rogers (1990) or Bergin (2000). In contrast, Bergin (2002) and Duarte and Wolman (2002) model currency unions allowing consumer price levels to differ across regions.

[16] The law of one price states that, absent trade barriers, a commodity should sell for the same price (when expressed in the same currency) everywhere.

[17] Of course, if the weight α differs across regions, then the price level will also differ across regions due to the difference in composition of the indices (even if p^N and p^T are identical across regions).

[18] Denoting the inflation rate in period t, the percentage change in the price level from $t-1$ to t, as π_t, it follows that the inflation differential can be approximated by a weighted average of the inflation differential in traded- and nontraded-goods indices:

$$\pi_{A,t} - \pi_{B,t} \simeq \alpha\left(\pi_{A,t}^{N} - \pi_{B,t}^{N}\right) + (1-\alpha)\left(\pi_{A,t}^{T} - \pi_{B,t}^{T}\right).$$

the sectors producing nontraded goods) than the other countries in the currency union, then this country will experience higher inflation than the other countries. To see this, note that a positive shock to productivity in the traded-goods sector leads to a higher real wage in the country (since labor is assumed to be perfectly mobile across sectors). In the nontraded-goods sector, the higher real wage drives up the relative price of nontraded goods, since productivity in this sector has not risen. Assuming that the law of one price holds for traded goods, a higher relative price of nontraded goods in this country raises this country's price level relative to that abroad. The inflation differential associated with the shock to productivity in the traded-goods sector is an equilibrium phenomenon and will persist while productivity differentials persist across countries. An inflation differential generated in this way is known as the Balassa-Samuelson effect.[19]

At the inception of a currency union, another source of inflation differentials is price level convergence. If price levels differ initially across countries, then adopting a common currency will lead prices to converge (at least to some extent), generating temporary inflation differentials across countries. Price level convergence can occur for both traded and nontraded goods.

For traded goods, increased market integration and price transparency associated with the adoption of a common currency reduces the scope for deviations from the law of one price, leading to temporary inflation differentials for traded-goods price indices.[20] As for the price of nontraded goods, the Balassa-Samuelson hypothesis also suggests that adopting a common currency narrows deviations in the price for these goods across countries. To see this, note that in a currency union, economic integration creates pressure for convergence in productivity levels. Since tradable goods tend to be more capital intensive than nontraded goods, the scope for productivity differentials across countries in the nontraded-goods sector tends to be limited relative to that in the traded-goods sector. Therefore, countries with initially low productivity levels (which tend to be poorer and have lower price levels) tend to experience higher productivity growth in the tradable-goods sector as a result of convergence in productivity. As we have seen before, prices of nontraded goods in the countries with lower price levels tend to increase, converging to the higher price level of wealthier countries.[21]

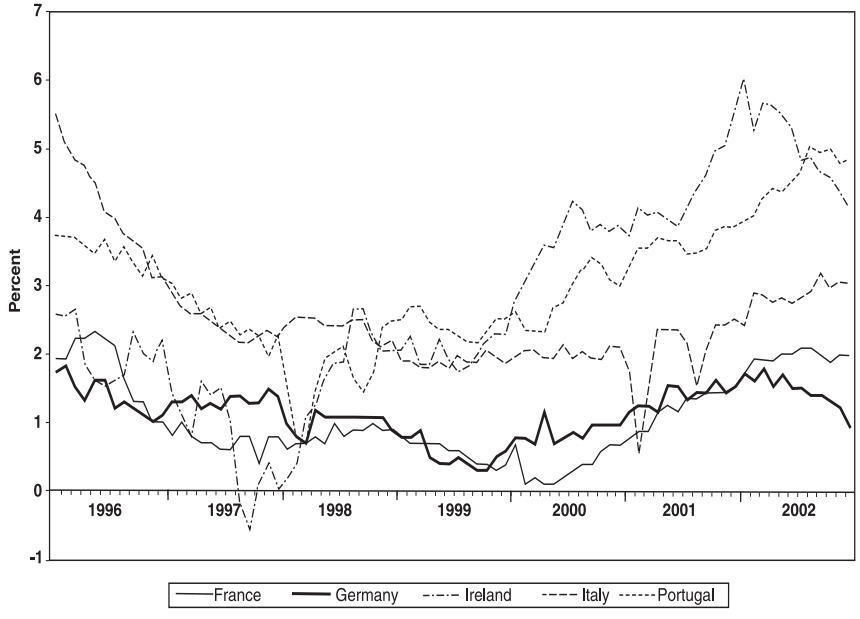
---

[19] See, for example, Chapter 4 in Obstfeld and Rogoff (1996) on the Balassa-Samuelson effect.

[20] The ECB *Monthly Bulletin* (October 1999), for example, provides strong evidence for the convergence of car prices across Euro-area countries.

[21] Natalucci and Ravenna (2002) analyze the choice of the exchange rate regime for accession countries to the EMU when these countries need to meet both inflation and nominal exchange rate criteria but are experiencing a real exchange rate appreciation due to increased productivity in the tradable-goods sector (the Balassa-Samuelson effect).

Figure 1 Monthly Core CPI Inflation



3. THE ADOPTION OF THE EURO AND THE BEHAVIOR OF INFLATION

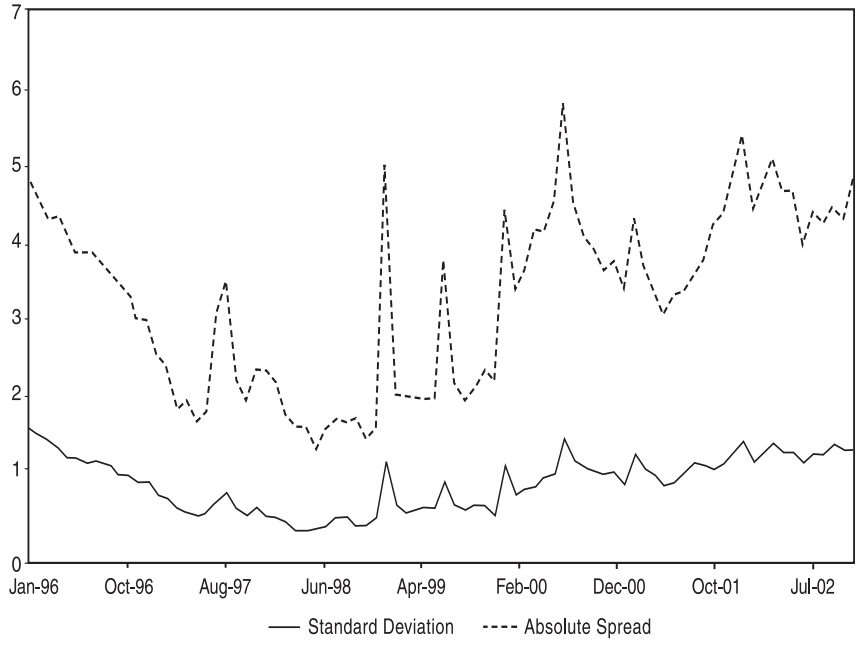
In this section I document the recent behavior of inflation in Euro-area countries. The measure of inflation I use is the twelve-month percentage change of the core consumer price index (CPI) for each of the eleven countries that adopted the Euro in 1999, at a monthly frequency.[22] That is, inflation in month t in country j is measured by $\pi_t^j = \frac{CPI_t^j}{CPI_{t-12}^j} - 1$.

Figure 1 depicts the monthly core consumer price inflation for a subset of Euro-area countries. Consumer price inflation declined steadily during the second half of the 1990s but has recently started to rise throughout the Euro zone. In the beginning of 1999, the twelve-month inflation rate in most countries was below the ECB's medium-term price stability target of 2 percent; this rate was above this level only in Portugal and Spain (2.5 percent). By mid-2002, the overall picture was quite different. Except for Germany and France,

[22] Core consumer price indices exclude food and energy prices, which are considered the most erratic components of price indices. The data are taken from Eurostat.

I have not included data for Greece, which adopted the Euro in January 2001.

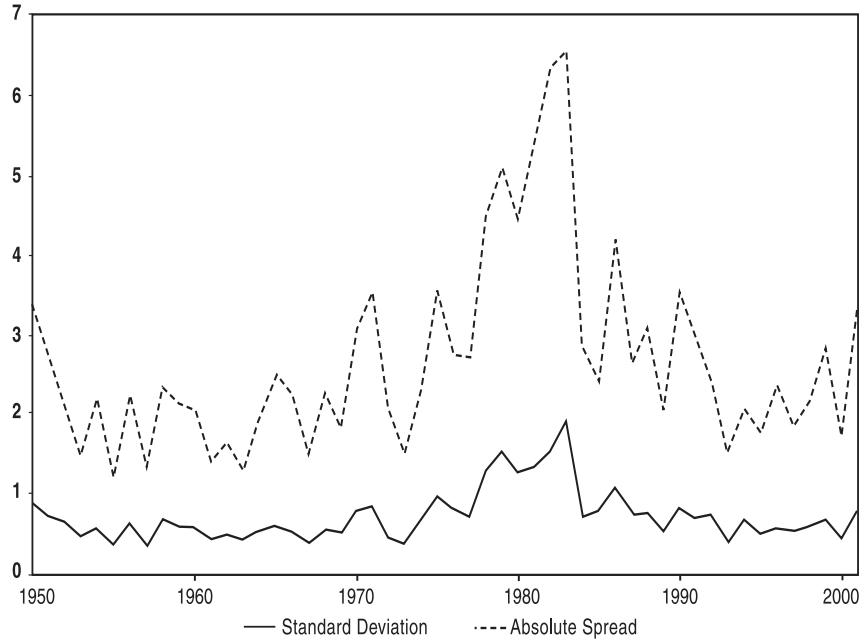
Figure 2 Euro-Area Inflation Dispersion: Absolute Spread and Standard Deviation



all Euro-zone countries were above the 2 percent target. The twelve-month core CPI inflation rate in June 2002 was 5.3 percent in Ireland, 4.5 percent in Portugal, and 3.9 percent in the Netherlands and Spain, for example.

I now turn to the behavior of inflation dispersion in the Euro area in this period. Figure 2 plots two summary statistics for the dispersion of inflation: the absolute difference between the highest and lowest inflation rates and the (unweighted) standard deviation of inflation rates across the Euro area from 1996:1 to 2002:12. The absolute spread decreased sharply during the second half of the 1990s, from about 4 percentage points to about 2 percentage points by the end of the decade, as the EU member countries aimed at fulfilling the convergence criteria defined by the Maastricht Treaty. The absolute spread has increased since then to nearly its level at the beginning of the sample (the average absolute spread in 2002 was 3.8 percentage points). The graph also shows a decrease in the standard deviation of inflation rates across the Euro area before the common currency was adopted followed by a subsequent increase. In 2002, the standard deviation averaged 1.2 percent, while in 1998 it averaged 0.6 percent.

Figure 3 U.S. Inflation Dispersion: Absolute Spread and Standard Deviation

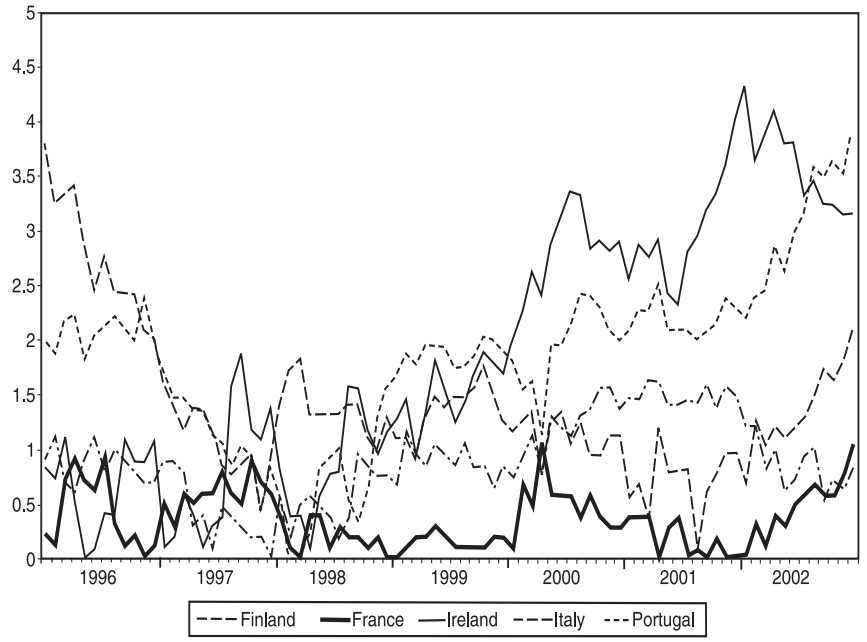


The United States constitutes a long-established currency union, and data on U.S. inflation dispersion provide a natural benchmark against which to compare the recent increase in inflation dispersion in Europe depicted in Figure 2. There is, however, relatively little data on U.S. subnational inflation rates. In order to compare inflation dispersion in the Euro area with that in the United States, I use annual data on consumer price levels in nineteen U.S. cities from the Bureau of Labor Statistics.

Figure 3 plots the two measures of inflation dispersion for the United States from 1950 to 2001. The average absolute spread was 2.8 percentage points in the entire sample and 2.2 percentage points in the last decade. The average standard deviation was 0.8 and 0.6 in these two periods, respectively.[23] Comparing the dispersion of inflation rates in the Euro area with that observed among U.S. cities indicates that the former resembled the latter in the late 1990s. Notwithstanding the existence of some episodes of high

[23] Cecchetti, Mark, and Sonora (2002) study the dynamics of these price indices for U.S. cities. They estimate that price index divergences across U.S. cities are temporary but surprisingly persistent, with a half-life of nearly nine years.

Figure 4 Absolute Inflation Differentials Relative to Germany



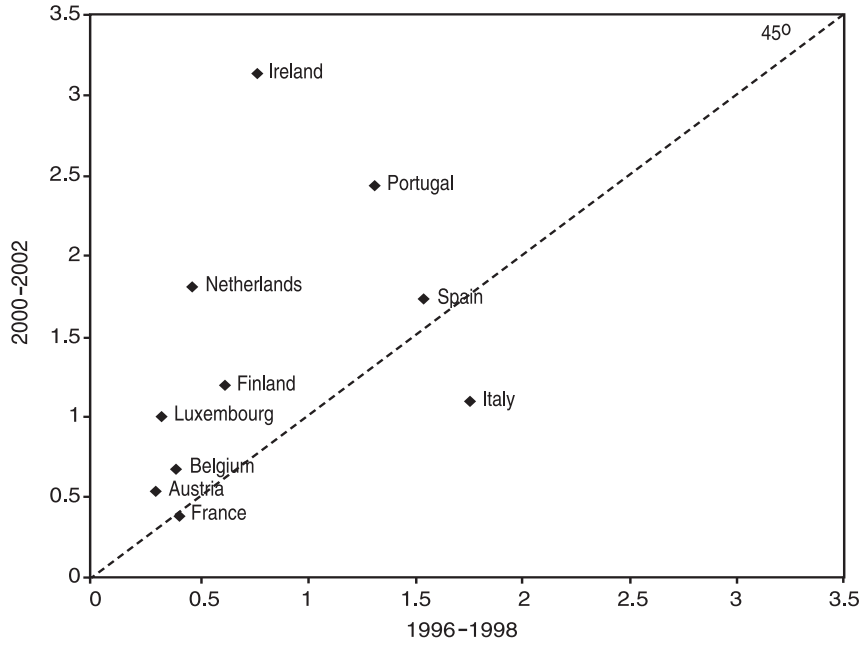
inflation dispersion in the United States, the two measures of Euro-area inflation dispersion are currently higher than the corresponding U.S. sample averages.[24]

I now turn from these two summary statistics of inflation dispersion to the distribution of inflation differentials with respect to the German inflation rate across the Euro area. The choice of Germany as the reference country reflects the fact that, prior to the adoption of the Euro, the German monetary authority had a strong reputation for advocating low inflation and that its inflation rate has been relatively flat throughout the period considered (Figure 1). Focusing on the distribution of inflation differentials allows some insight into the nature of these differentials.

Figure 4 plots the inflation differential with respect to German inflation for a subset of Euro-zone countries, and Figure 5 plots the average inflation differential with respect to German inflation for each Euro-area member country before and after the adoption of the Euro. The former period averages inflation data from 1996 through 1998, and the latter period averages inflation data from 2000 onwards. I have not included the twelve data points for 1999 since

[24] In comparing Figures 2 and 3, the distinct time samples as well as the distinct frequency of the data should be noted. Using annual (instead of monthly) frequency data for the EMU countries leads to the same conclusion.

Figure 5 Average Inflation Differentials (percentage points)



the twelve-month percentage change of consumer price indices for these data points effectively cover both the period before and after the Euro was adopted.

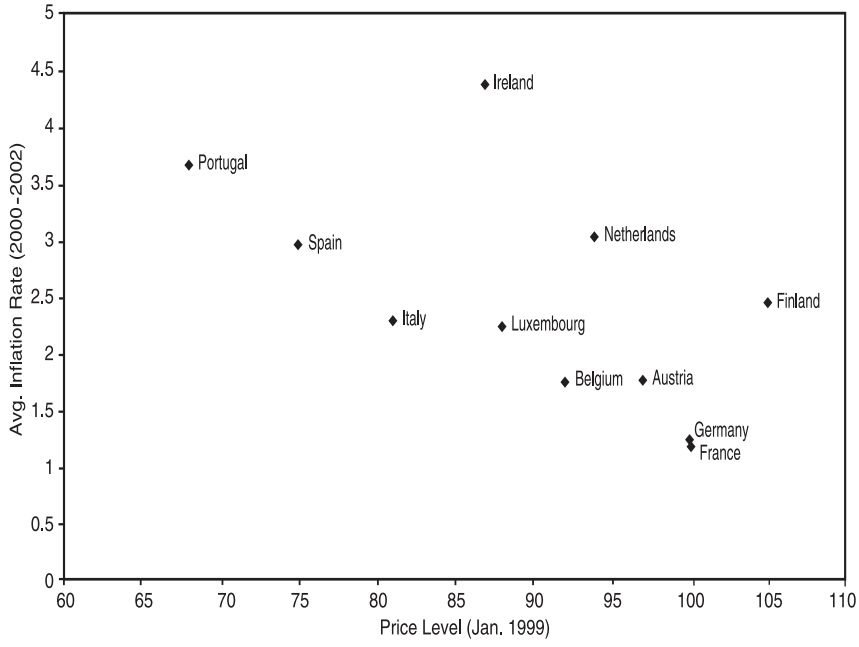
It is apparent from these two figures that inflation differentials within the Euro area have increased after the adoption of the common currency for most countries, reinforcing the message from Figure 2. Over the period before the Euro was adopted (1996 to 1998), average inflation differentials ranged from 0.3 (Austria) to 1.7 (Italy) percentage points. Inflation differentials have increased steadily across the Euro since after 1999, and over the period from 2000 to 2002 they ranged from 0.4 (France) to more than 3 (Ireland) percentage points.

Assessing the sources of the inflation differentials observed in the Euro area after 1999 is a complicated task. Drawing upon the discussion in the previous section, I briefly look at the joint behavior of inflation with the growth rate of output and initial price levels.

As I pointed out in the previous section, price level convergence can be a source of inflation differentials when different countries with initial distinct price levels adopt a common currency.[25] This argument suggests that countries

[25] The ECB has emphasized price level convergence as an important source of inflation differentials in the Euro area. See, for example, the ECB *Monthly Bulletin* (October 1999).

Figure 6 Initial Price Levels and Average Inflation Rate



with lower price levels would exhibit higher inflation rates than countries with higher price levels. I use the comparative price levels from the OECD for January 1999 as a measure of the initial differences in price levels among the countries in the Euro zone. Figure 6 plots the average inflation rate after 2000 against the comparative price level in 1999 for each country in the Euro area. The plot shows a negative relationship between the price level and subsequent inflation rates (the correlation coefficient is -0.6).[26] This evidence suggests that price level convergence may be a partial explanation for the different behavior of inflation across Euro-zone countries. The process of price level convergence, however, is a temporary one, and it has been under way in Europe throughout the 1990s.[27] This fact suggests a reduced scope for future price level convergence in Europe.

[26] Comparative price levels, a measure of the differences in price levels between countries, are from the OECD Main Economic Indicators.

[27] See Rogers (2001) for evidence on price level convergence in Europe during the 1990s. He concludes that while price level convergence contributed to observed inflation differentials within the Euro area in 2000, other forces explain most of those cross-country differences in inflation.

Figure 7 Average Inflation Rates and GDP Growth Rates After 1999

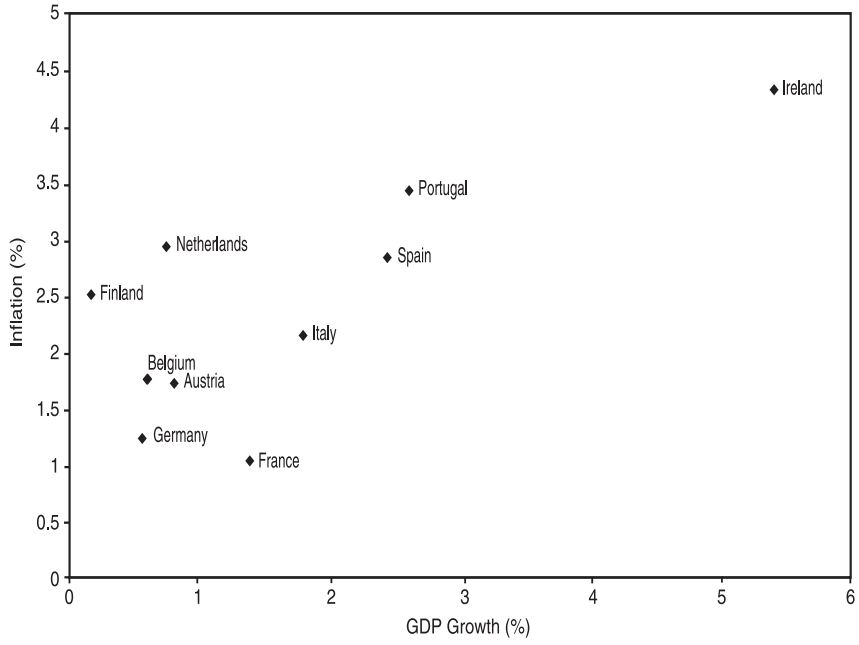
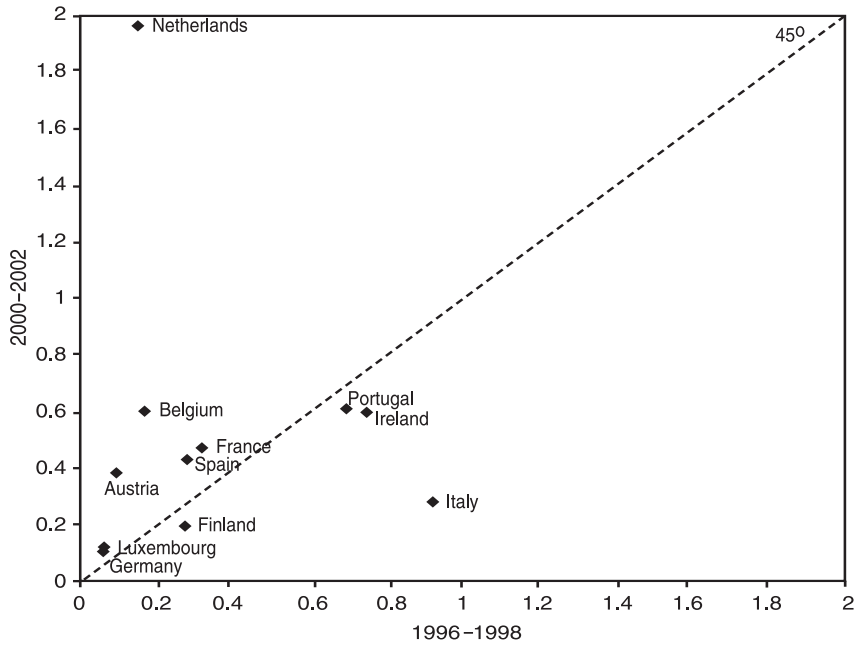


Figure 7 plots the average inflation rate after the Euro was adopted in each Euro-zone country against its average growth rate of GDP in the same period. This figure clearly suggests a positive relationship between the average growth rate of output and average inflation after the common currency was adopted. This figure suggests that, reflecting the Balassa-Samuelson hypothesis, the observed inflation differentials could be indicative of a process of the convergence of productivity levels (driving income convergence) across countries as well as asymmetric shocks across countries (and desynchronized business cycles).

Finally, in Figure 8, I plot the variance of twelve-month inflation in the periods before and after the adoption of the Euro, as defined before. This figure shows a tendency for increased inflation variability after the adoption of the Euro relative to the previous period. The variance of inflation increased in the later period for seven out of the eleven countries considered. The most significant exception is Italy, where the variance of inflation was substantially smaller after the Euro was adopted.

The Maastricht criteria forced the potential entrants in the EMU to attain inflation convergence by 1998, a requirement that conditioned these countries' use of monetary and fiscal policy throughout the 1990s. With the start of the

Figure 8 Variance of Inflation



EMU, the restriction on inflation convergence was eliminated and the ECB took control of monetary policy in the Euro area. The figures above suggest that the inability to use monetary policy in response to country-specific shocks after the requirement that countries attain inflation convergence was eliminated is associated with an increase of inflation dispersion and volatility.

In the new European institutional framework, regional fiscal policy is the only instrument available to the regional authorities to affect regional inflation. Should a regional fiscal authority decide to use fiscal policy to affect its inflation rate, it raises the question of the effectiveness and implications of such policy.[28]

4. CONCLUSION

In this article I document the behavior of inflation dispersion and inflation differentials in Euro-area countries before and after the Euro was introduced. Inflation dispersion and inflation differentials (with respect to German inflation)

[28] See Duarte and Wolman (2003) for an analysis of the implications of using fiscal policy to affect regional inflation in a currency union.

within the Euro area have increased since countries lost monetary independence and were no longer required to attain inflation convergence. Inflation dispersion in the Euro area after the common currency was adopted has been higher than that observed in the United States. Additionally, the variability of the inflation differential with respect to German inflation has tended to increase for most countries since the Euro was adopted.

These observed inflation differentials reflect both a temporary process of price level convergence as well as the adjustment to asymmetric country-specific shocks. To the extent that the process of price level convergence is temporary, these differentials, if continued or widened, are bound to start generating considerable attention, prompting the debate over the criteria to be met by Euro-area countries and the design and goals of regional policies. These differentials naturally raise the question of the adequacy of a common monetary policy for an area composed of heterogeneous constituent countries and, since fiscal policy is the only tool available to regional authorities to affect inflation, the question of the ability and desirability of using regional fiscal policy to affect regional price differentials.

REFERENCES

Bergin, Paul. 2000. "Fiscal Solvency and Price Level Determination in a Monetary Union." *Journal of Monetary Economics* 25 (February): 37–53.

_____. Forthcoming. "One Money One Price? Pricing to Market in a Monetary Union." *European Economic Review*.

Canzoneri, Matthew, and Caroll Ann Rogers. 1990. "Is the European Community an Optimal Currency Area? Optimal Taxation Versus the Cost of Multiple Currencies." *American Economic Review* 80 (June): 419–33.

Cecchetti, Stephen, Nelson Mark, and Robert Sonora. 2002. "Price Level Convergence Among United States Cities: Lessons for the European Central Bank." *International Economic Review* 43 (November): 1081–99.

Chari, V. V., Patrick Kehoe, and Ellen McGrattan. Forthcoming. "Can Sticky Price Models Generate Volatile and Persistent Exchange Rates?" *Review of Economic Studies*.

Duarte, Margarida, and Alexander Wolman. 2003. "Fiscal Policy and Regional Inflation in a Currency Union." Mimeo, Federal Reserve Bank

of Richmond.

European Central Bank. 1999. "Inflation Differentials in a Monetary Union." *Monthly Bulletin* (October): 35–44.

Mundell, Robert. 1961. "A Theory of Optimum Currency Areas." *American Economic Review* 51 (September): 657–65.

Natalucci, Fabio, and Federico Ravenna. 2002. "The Road to Adopting the Euro: Monetary Policy and Exchange Rate Regimes in EU Candidate Countries." Mimeo, Federal Reserve Board of Governors.

Obstfeld, Maurice, and Kenneth Rogoff. 1995. "Exchange Rate Redux." *Journal of Political Economy* 103 (June): 624–60.

_____. 2002. "Risk and Exchange Rates." In *Contemporary Economic Policy: Essays in Honor of Assaf Razin*, edited by E. Helpman and E. Sadka. Cambridge: Cambridge University Press.

Parsley, David, and Shang-Jin Wei. 1996. "Convergence to the Law of One Price Without Trade Barriers or Currency Fluctuations." *Quarterly Journal of Economics* 111 (November): 1211–36.

Rogers, John. 2001. "Price Level Convergence, Relative Prices, and Inflation in Europe." Mimeo, Federal Reserve Board of Governors.

Rose, Andrew. 2002. "The Effect of Common Currencies on International Trade: A Meta-Analysis." Mimeo, Haas School of Business, University of California–Berkeley.

_____, and Eric van Wincoop. 2001. "National Money as a Barrier to Trade: The Real Case for Currency Union." *American Economic Review* 91 (May): 386–90.

Sibert, Anne. 1994. "The Allocation of Seigniorage in a Common Currency Area." *Journal of International Economics* 37 (August): 111–22.

Svensson, Lars. 2002. "A Reform of the Eurosystem's Monetary-Policy Strategy Is Increasingly Urgent." Mimeo, Princeton University.

How Did Leading Indicator Forecasts Perform During the 2001 Recession?

James H. Stock and Mark W. Watson

The recession that began in March 2001 differed in many ways from other recessions of the past three decades. The twin recessions of the early 1980s occurred when the Federal Reserve Board, under Chairman Paul Volcker, acted decisively to halt the steady rise of inflation during the 1970s, despite the substantial employment and output cost to the economy. Although monetary tightening had reduced the growth rate of real activity in 1989, the proximate cause of the recession of 1990 was a sharp fall in consumption, a response by consumers to the uncertainty raised by Iraq's invasion of Kuwait and the associated spike in oil prices (Blanchard 1993). In contrast, the recession of 2001 started neither in the shopping mall nor in the corridors of the Federal Reserve Bank, but in the boardrooms of corporate America as businesses sharply cut back on expenditures—most notably, investment associated with information technology—in turn leading to declines in manufacturing output and in the overall stock market.

Because it differed so from its recent predecessors, the recession of 2001 provides a particularly interesting case in which to examine the forecasting performance of various leading economic indicators. In this article, we take a look at how a wide range of leading economic indicators performed during this episode. Did these leading economic indicators predict a slowdown of growth? Was that slowdown large enough to suggest that a recession was

James H. Stock is with the Department of Economics at Harvard University and the National Bureau of Economic Research. Mark W. Watson is with the Woodrow Wilson School and Department of Economics at Princeton University and the National Bureau of Economic Research. The authors would like to thank Frank Diebold, Marvin Goodfriend, Yash Mehra, and Roy Webb for helpful comments on an earlier draft of this article. The views expressed in this article are not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

imminent? Were the leading indicators that were useful in earlier recessions also useful in this recession? Why or why not?

We begin our analysis by examining the predictions of professional forecasters—specifically, the forecasters in the Survey of Professional Forecasters (SPF) conducted by the Federal Reserve Bank of Philadelphia—during this episode. As we show in Section 2, these forecasters were taken by surprise: even as late as the fourth quarter of 2000, when industrial production was already declining, the median SPF forecast was predicting strong economic growth throughout 2001.

Against this sobering backdrop, Section 3 turns to the performance of individual leading indicators before and during the 2001 recession. Generally speaking, we find that the performance of specific indicators was different during this recession. Some indicators, in particular the so-called term spread (the difference between long-term and short-term interest rates on government debt) and stock returns, provided some warning of a slowdown in economic growth, although the predicted growth was still positive and these indicators fell short of providing a signal of an upcoming recession. Other, previously reliable leading indicators, such as housing starts and orders for capital goods, provided little or no indication of the slowdown.

In practice, individual leading indicators are not used in isolation; as Mitchell and Burns (1938) emphasized when they developed the system of leading economic indicators, their signals should be interpreted collectively. Accordingly, Section 4 looks at the performance of pooled forecasts based on the individual leading indicator forecasts from Section 3 and finds some encouraging results. Section 5 concludes.

1. FORECASTING THE 2001 RECESSION: HOW DID THE PROS DO?

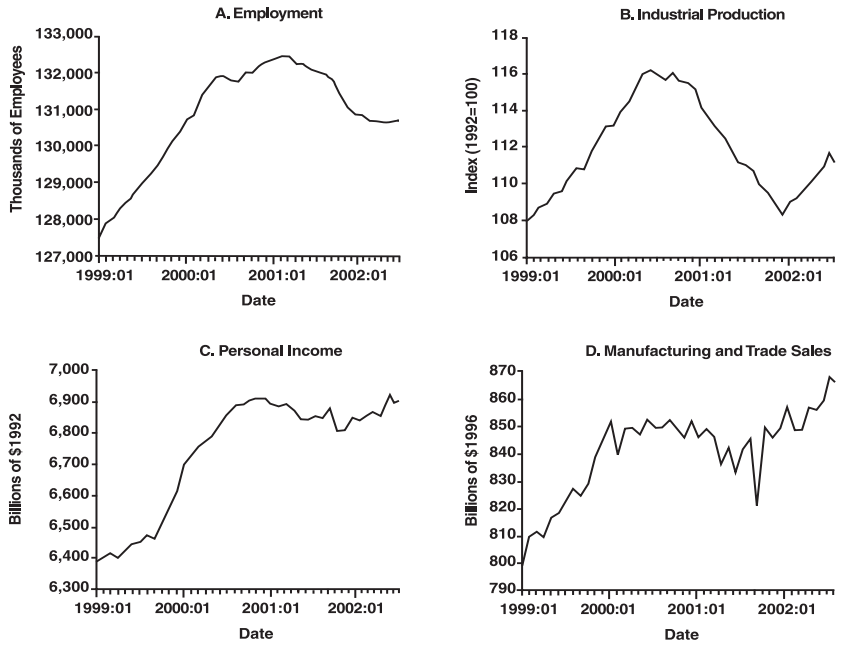
This section begins with a brief quantitative review of the 2001 recession. We then turn to professional forecasts during this episode, as measured in real time by the Philadelphia Fed's quarterly Survey of Professional Forecasters.

A Brief Reprise of the 2001 Recession

Figure 1 presents monthly values of the four coincident indicators that constitute the Conference Board's Index of Coincident Indicators: employment in nonagricultural businesses, industrial production, real personal income less transfers, and real manufacturing and trade sales.[1] These four series are also the primary series that the NBER Business Cycle Dating Committee uses to

[1] For additional information on the Conference Board's coincident and leading indexes, see www.tcb-indicators.org.

Figure 1 Coincident Indicators

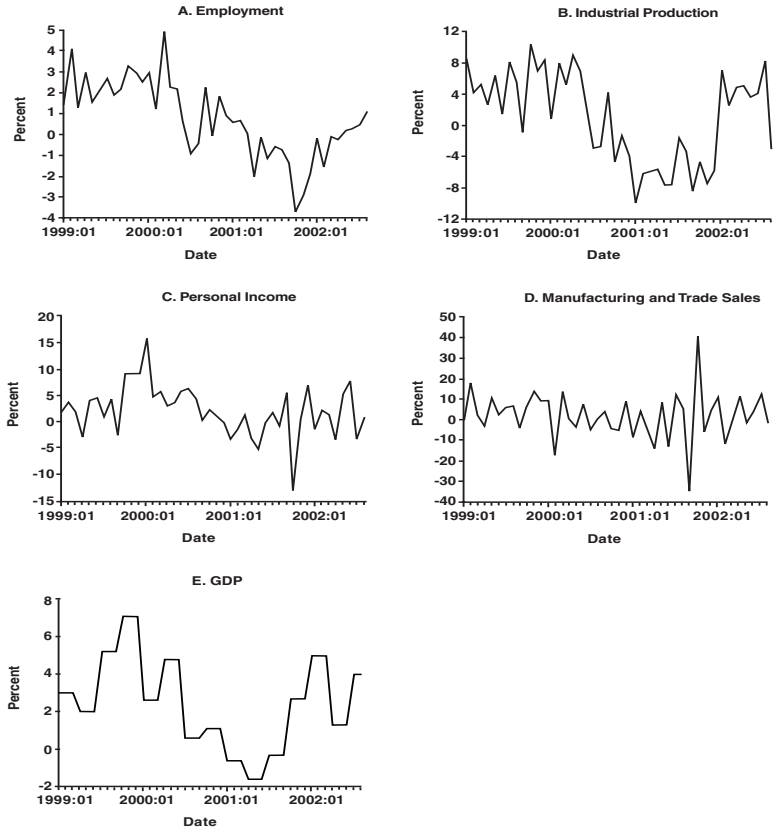


establish its business cycle chronology (Hall 2002). The percentage growth rates of these series, expressed at an annual rate, are plotted in Figure 2. In addition, Figure 2 presents the percentage growth of real GDP (at an annual rate); because GDP is measured quarterly and the time scale of Figure 2 is monthly, in Figure 2 the same growth rate of real GDP is attributed to each month in the quarter, accounting for the "steps" in this plot.

Figures 1 and 2 reveal that the economic slowdown began with a decline in industrial production, which peaked in June 2000. Manufacturing and trade sales fell during the first quarter of 2001, but employment did not peak until March 2001, the official NBER cyclical peak. Real personal income reached a cyclical peak in November 2000 and declined by 1.5 percent over the next twelve months. This relatively small decline in personal income reflected the unusual fact that productivity growth remained strong through this recession. Based on the most recently available data, real GDP fell during the first three quarters of 2001, with a substantial decline of 1.6 percent (at an annual rate) in the second quarter.

The economy gained substantial strength in the final quarter of 2001 and throughout 2002, and all the monthly indicators were growing by December 2001. Thus, based on the currently available evidence, the recession appears

Figure 2 Coincident Indicators (Growth Rates, PAAR)



to have ended in the fourth quarter of 2001. When this article went into production, however, the NBER had yet to announce a cyclical trough, that is, a formal end to the recession.

Professional Forecasts During 2000 and 2001

In the second month of every quarter, the Research Department of the Federal Reserve Bank of Philadelphia surveys a large number of professional forecasters—in the first quarter of 2000, thirty-six forecasters or forecasting groups participated—and asks them a variety of questions concerning their short-term forecasts for the U.S. economy. Here, we focus on two sets of forecasts: the forecast of the growth rate of real GDP, by quarter, and the

Table 1 Median Forecasts of the Percentage Growth in Quarterly GDP from the Survey of Professional Forecasters

Target Date		Forecasts Made In							
		2000				2001			
Quarter	**Actual Growth**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
2000Q4	1.1	2.9	3.1	**3.2**	3.2				
2001Q1	−0.6	2.8	2.6	3.0	**3.3**	0.8			
2001Q2	−1.6		2.9	2.7	3.2	**2.2**	1.2		
2001Q3	−0.3			3.2	3.3	3.3	**2.0**	1.2	
2001Q4	2.7				3.2	3.7	2.6	**2.8**	−1.9
2002Q1	5.0					3.7	3.1	2.7	**0.1**
2002Q2	1.3						3.6	3.0	2.4
2002Q3	4.0							3.9	3.6

Notes: Entries are quarterly percentage growth rates of real GDP, at an annual rate. One-quarter-ahead forecasts appear in bold. Actual GDP growth is from the 28 February 2003 GDP release by the Bureau of Economic Analysis. Forecasts are the median forecast from the Philadelphia Federal Reserve Bank's Survey of Professional Forecasters (various issues; see www.phil.frb.org/econ/spf).

probability that the forecasters assign to the event that GDP growth will be negative in an upcoming quarter.

The median growth forecasts—that is, the median of the SPF panel of forecasts of real GDP growth for a given quarter—are summarized in Table 1 for late 2000Q4 through 2002Q3. The first two columns of Table 1 report the quarter being forecast and its actual growth rate of real GDP, based on the most recently available data as of this writing. The remaining columns report the median SPF growth forecasts; the column date is the quarter in which the forecast is made for the quarter of the relevant row. For example, as of 2000Q1, the SPF forecast for 2000Q4 GDP growth was 2.9 percent at an annual rate (this is the upper-left forecast entry in Table 1). Over the course of 2000, as the fourth quarter approached, the SPF forecast of 2000Q4 growth rose slightly; as of 2000Q3, the forecast was 3.2 percent. Because the Bureau of Economic Analysis does not release GDP estimates until the quarter is over, forecasters do not know GDP growth for the current quarter, and in the 2000Q4 survey the average SPF forecast of 2000Q4 real GDP growth was 3.2 percent. As it happened, the actual growth rate of real GDP during that quarter was substantially less than forecasted, only 1.1 percent based on the most recently available data.

An examination of the one-quarter-ahead forecasts (for example, the 2000Q3 forecast of 2000Q4 growth) and the current-quarter forecasts (the 2000Q4 forecast of 2000Q4 growth) reveals that the SPF forecasters failed

Table 2 Probabilities of a Quarterly Decline in Real GDP from the Survey of Professional Forecasters

Target Date		Forecasts Made In							
		2000				2001			
Quarter	Actual Growth	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
2000Q4	1.1	13%	9%	**7%**	4%				
2001Q1	−0.6	17	15	13	**11**	37%			
2001Q2	−1.6		18	16	17	**32**	32%		
2001Q3	−0.3			17	19	23	**29**	35%	
2001Q4	2.7				19	18	23	**26**	82%
2002Q1	5.0					13	18	20	**49**
2002Q2	1.3						13	16	27
2002Q3	4.0							15	18

Notes: Forecast entries are the probability that real GDP growth will be negative, averaged across SPF forecasters. The forecasted probability that growth will be negative in the quarter after the forecast is made (that is, the one-quarter-ahead forecast) appears in bold. See the notes to Table 1.

to predict the sharp declines in real GDP, even as they were occurring. The SPF one-quarter-ahead forecast of 2001Q1 growth was 3.3 percent, whereas GDP actually fell by 0.6 percent; the one-quarter-ahead forecast of 2001Q2 growth was 2.2 percent, but GDP fell by 1.6 percent; and the one-quarter-ahead forecast of 2001Q3 growth was 2.0 percent, while GDP fell by 0.3 percent. Throughout this episode, this average forecast was substantially too optimistic about near-term economic growth. Only in the fourth quarter of 2001 did the forecasters begin to forecast ongoing weakness—in part in reaction to the events of September 11—but, as it happened, in that quarter GDP was already recovering.

The SPF forecasters are also asked the probability that real GDP will fall, by quarter, and Table 2 reports the average of these probabilities across the SPF forecasters. In the fourth quarter of 2000, the forecasters saw only an 11 percent chance that GDP growth in the first quarter of 2001 would be negative, consistent with their optimistic growth forecast of 3.3 percent for that quarter; in fact, GDP growth was negative, falling by 0.6 percent. Throughout the first three quarters of 2001, the current-quarter predicted probabilities of negative growth hovered around one-third, even though growth was in fact negative in each of those quarters. When, in the fourth quarter of 2001, the SPF forecasters finally were sure that growth would be negative—the SPF probability of negative same-quarter growth was 82 percent—the economy in fact grew by a strong 2.7 percent. Evidently, this recession was a challenging time for professional forecasters.

Table 3 Relative MSFEs of Individual Indicator Forecasts of U.S. Output Growth, 1999Q1–2002Q3

Predictor	Transformation	GDP		IP	
		$h = 2$	$h = 4$	$h = 2$	$h = 4$
		Root Mean Squared Forecast Error			
Univariate autoregression		2.06	2.03	4.34	4.92
Predictor		*MSFE Rel. to Univariate AR Model*			
Random walk	level	1.26	1.11	1.56	1.17
Interest Rates					
Federal funds	Δ	1.01	0.71	0.97	0.78
90-day T-bill	Δ	1.01	0.76	1.02	0.88
1-year T-bond	Δ	1.17	0.96	1.22	1.06
5-year T-bond	Δ	1.37	1.24	1.38	1.23
10-year T-bond	Δ	1.36	1.26	1.21	1.23
Spreads					
Term spread (10 year–federal funds)*	level	0.86	0.65	0.77	0.72
Term spread (10 year–90-day T-bill)	level	0.87	0.62	0.70	0.62
Paper-bill spread (commercial paper–T-bill)	level	1.31	1.17	1.96	1.43
Junk bond spread (high yield–AAA corporate)	level	0.76	0.65	0.67	0.58
Other Financial Variables					
Exchange rate	Δ ln	0.85	0.87	0.85	0.80
Stock prices*	Δ ln	0.83	0.93	0.64	0.71
Output					
Real GDP	Δ ln			0.92	0.96
IP–total	Δ ln	0.98	1.01		
IP–products	Δ ln	1.03	0.99	1.03	0.96
IP–business equipment	Δ ln	1.00	1.01	1.05	1.06
IP–intermediate products	Δ ln	0.89	0.90	0.89	0.88
IP–materials	Δ ln	0.97	1.01	1.04	0.98
Capacity utilization rate	level	0.91	1.01	0.85	1.03
Labor Market					
Employment	Δ ln	0.96	1.00	0.96	0.99
Unemployment rate	Δ	1.24	1.08	1.31	1.09
Average weekly hours in manufacturing*	level	0.87	0.75	0.72	0.87
New claims for unemployment insurance*	Δ ln	0.75	0.84	0.74	0.81

Continued on next page

2. FORECASTS BASED ON INDIVIDUAL LEADING INDICATORS

Perhaps one reason for these difficulties was that the 2001 recession differed from its recent predecessors. If so, this difference would also be reflected

Table 3 Relative MSFEs of Individual Indicator Forecasts of U.S. Output Growth, 1999Q1–2002Q3

Other Leading Indicators					
Housing starts (building permits)*	$\Delta \ln$	1.30	1.07	1.52	1.14
Vendor performance*	level	1.02	0.97	1.19	0.97
Orders–consumer goods and materials*	$\Delta \ln$	0.77	0.83	0.81	0.83
Orders–nondefense capital goods*	$\Delta \ln$	1.02	1.03	0.92	1.09
Consumer expectations (Michigan)*	level	1.96	2.14	1.33	1.49
Prices and Wages					
GDP deflator	$\Delta^2 \ln$	1.00	0.94	0.94	0.84
PCE deflator	$\Delta^2 \ln$	1.01	1.05	0.99	0.99
PPI	$\Delta^2 \ln$	1.01	1.02	0.96	0.99
Earnings	$\Delta^2 \ln$	1.00	1.01	0.89	0.98
Real oil price	$\Delta^2 \ln$	1.13	1.18	1.07	1.11
Real commodity price	$\Delta^2 \ln$	1.04	1.00	1.12	1.09
Money					
Real M0	$\Delta \ln$	2.13	2.84	1.41	1.73
Real M1	$\Delta \ln$	1.09	1.07	1.57	1.12
Real M2*	$\Delta \ln$	2.06	1.82	2.13	1.94
Real M3	$\Delta \ln$	1.81	2.23	2.05	2.15

Notes: The entry in the first line is the root MSFE of the AR forecast, in percentage growth rates at an annual rate. The remaining entries are the MSFE of the forecast based on the individual indicator, relative to the MSFE of the benchmark AR forecast. The first forecast is made using data through 1999Q1; the final forecast period ends at 2000Q3. The second column provides the transformation applied to the leading indicator to make the forecast, for example, for the federal funds rate forecasts, X_t in (1) is the first difference of the federal funds rate.

*Included in the Conference Board's Index of Leading Indicators.

in the performance of leading indicators over this episode. In this section, we examine the performance of forecasts based on individual leading indicators during the 2001 recession. We begin by discussing the methods used to construct these forecasts, then turn to graphical and quantitative analyses of the forecasts.

Construction of Leading Indicator Forecasts

The leading indicator forecasts were computed by regressing future output growth over two or four quarters against current and past values of output growth and the candidate leading indicator. Specifically, let $Y_t = \Delta \ln Q_t$, where Q_t is the level of output (either the level of real GDP or the Index of Industrial Production), and let X_t be a candidate predictor (e.g., the term

spread). Let Y^h_{t+h} denote output growth over the next h quarters, expressed at an annual rate; that is, let $Y^h_{t+h} = (400/h)\ln(Q_{t+h}/Q_t)$. The forecasts of Y^h_{t+h} are made using the h-step-ahead regression model,

$$Y^h_{t+h} = \alpha + \sum_{i=0}^{p-1} \beta_i X_{t-i} + \sum_{i=0}^{q-1} \gamma_i Y_{t-i} + u^h_{t+h}, \tag{1}$$

where u^h_{t+h} is an error term and $\alpha, \beta_0, \ldots, \beta_{p-1}, \gamma_0, \ldots, \gamma_{q-1}$ are unknown regression coefficients. Forecasts are computed for two- and four-quarter horizons ($h = 2$ and $h = 4$).

To simulate real-time forecasting, the coefficients of equation (1) were estimated using only data prior to the forecast date. For example, for a forecast made using data through the fourth quarter of 2000, we estimate (1) using only data available through the fourth quarter of 2000. Moreover, the number of lags of X and Y included in (1), that is, p and q, were also estimated using only data available through the date of the forecast; specifically, p and q were selected using the Akaike Information Criterion (AIC), with $1 \le p \le 4$ and $0 \le q \le 4$.[2] Restricting the estimation to data available through the forecast date—in this example, 2000Q4—prevents the forecasts from being misleadingly accurate by using future data and also helps to identify shifts in the forecasting relation during the period that matters for forecasting, the end of the sample. This approach, in which all estimation and model selection is done using only data prior to the forecast date, is commonly called "pseudo out-of-sample forecasting"; for an introduction to pseudo out-of-sample forecasting methods and examples, see Stock and Watson (2003b, Section 12.7).

As a benchmark, we computed a multistep autoregressive (AR) forecast, in which (1) is estimated with no X_t predictor and the lag length is chosen using the AIC ($0 \le q \le 4$). As an additional benchmark, we computed a recursive random walk forecast, in which $\hat{Y}^h_{t+h|t} = h\hat{\mu}_t$, where $\hat{\mu}_t$ is the sample average of Y_s, $s = 1, \ldots, t$. Like the leading indicator forecasts, these benchmark forecasts were computed following the pseudo out-of-sample methodology.[3]

[2] The AIC is $\mathrm{AIC}(p,q) = \ln(SSR_{p,q}/T) + 2(p+q+1)/T$, where $SSR_{p,q}$ is the sum of squared residuals from the estimation of (1) with lag lengths p and q, and T is the number of observations. The lag lengths p and q are chosen to minimize $\mathrm{AIC}(p,q)$ by trading off better fit (the first term) against a penalty for including more lags (the second term). For further explanation and a worked example, see Stock and Watson (2003b, Section 12.5).

[3] One way that this methodology does *not* simulate real-time forecasting is that we use the most recently available data to make the forecasts, rather than the data that were actually available in real time. For many of the leading indicators, such as interest rates and consumer expectations, the data are not revised, so this is not an issue. For others, such as GDP, revisions can be large, and because our simulated real-time forecasts use GDP growth as a predictor in equation (1), their performance in this exercise could appear better than it might have in real time, when preliminary values of GDP would be used.

A Look at Twelve Leading Indicators

We begin the empirical analysis by looking at the historical paths of twelve commonly used monthly leading indicators. After describing the twelve indicators, we see how they fared during the 2001 recession.

The Twelve Leading Indicators

Six of these indicators are based on interest rates or prices: a measure of the term spread (the ten-year Treasury bond rate minus the federal funds rate); the federal funds rate; the paper-bill spread (the three-month commercial paper rate minus the Treasury bill rate); the high-yield "junk" bond spread (the difference between the yield on high-yield securities[4] and the AAA corporate bond yield); the return on the S&P 500; and the real price of oil. Research in the late 1980s (Stock and Watson 1989; Harvey 1988, 1989; Estrella and Hardouvelis 1991) provided formal empirical evidence supporting the idea that an inverted yield curve signals a recession, and the term spread is now one of the seven indicators in the Conference Board's Index of Leading Indicators (ILI). The federal funds rate is included because it is the instrument of monetary policy. Public-private spreads also have been potent indicators in past recessions (Stock and Watson 1989; Friedman and Kuttner 1992); the second of these, the junk bond spread, was proposed by Gertler and Lown (2000) as an alternative to the paper-bill spread, which failed to move before the 1991 recession. Stock returns have been a key financial leading indicator since they were identified as such by Mitchell and Burns (1938), and the S&P 500 return is included in ILI.[5] Finally, fluctuations in oil prices are widely considered to be a potentially important source of external economic shocks and have been associated with past recessions (e.g., Hamilton 1983).

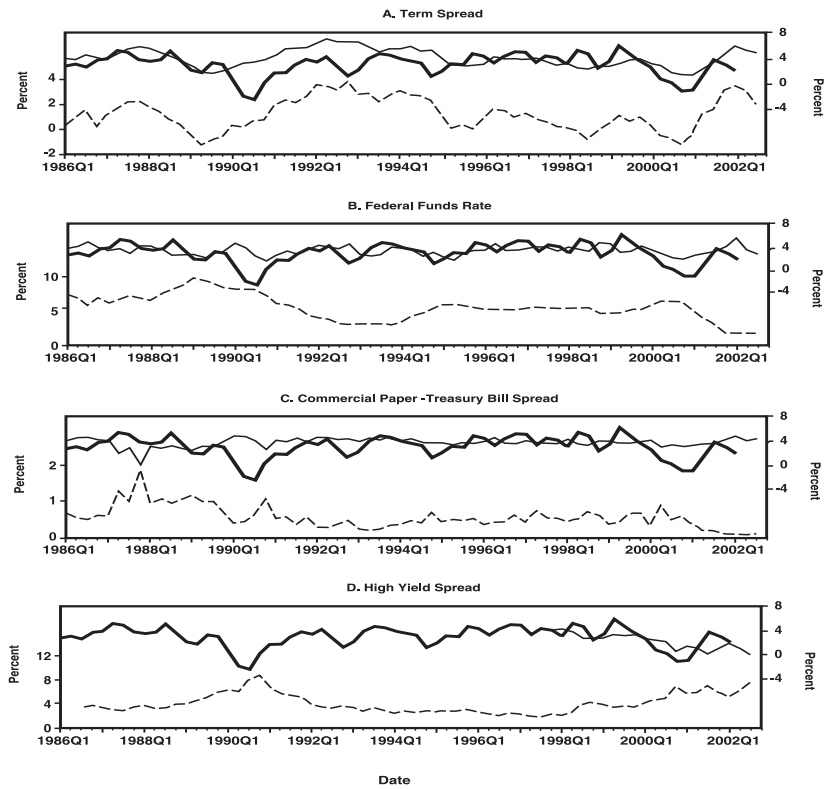
The next five indicators measure different aspects of the real economy. Three of these are in the ILI: new claims for unemployment insurance; housing starts (building permits); and the University of Michigan Index of Consumer Expectations. Because corporate investment played a central role in the 2001 recession, we also look at two broad monthly measures of business investment: industrial production of business equipment and new orders for capital goods. Finally, we consider a traditional leading indicator, the growth rate of real M2, which also enters the ILI.

---

[4] Merrill Lynch, U.S. High Yield Master II Index.

[5] For a review of the extensive literature over the past fifteen years on the historical and international performance of asset prices as leading indicators, see Stock and Watson (2001).

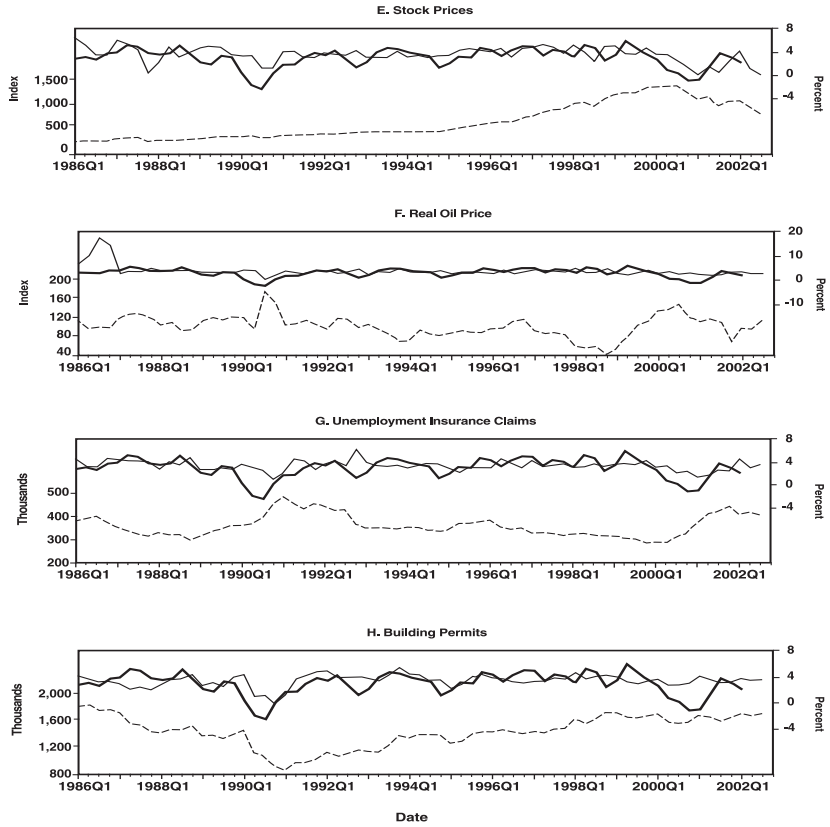
Figure 3 Twelve Leading Indicators from 1986 to 2002, Two-Quarter Growth in Real GDP, and Its Leading-Indicator-Based Forecast



Graphical Analysis

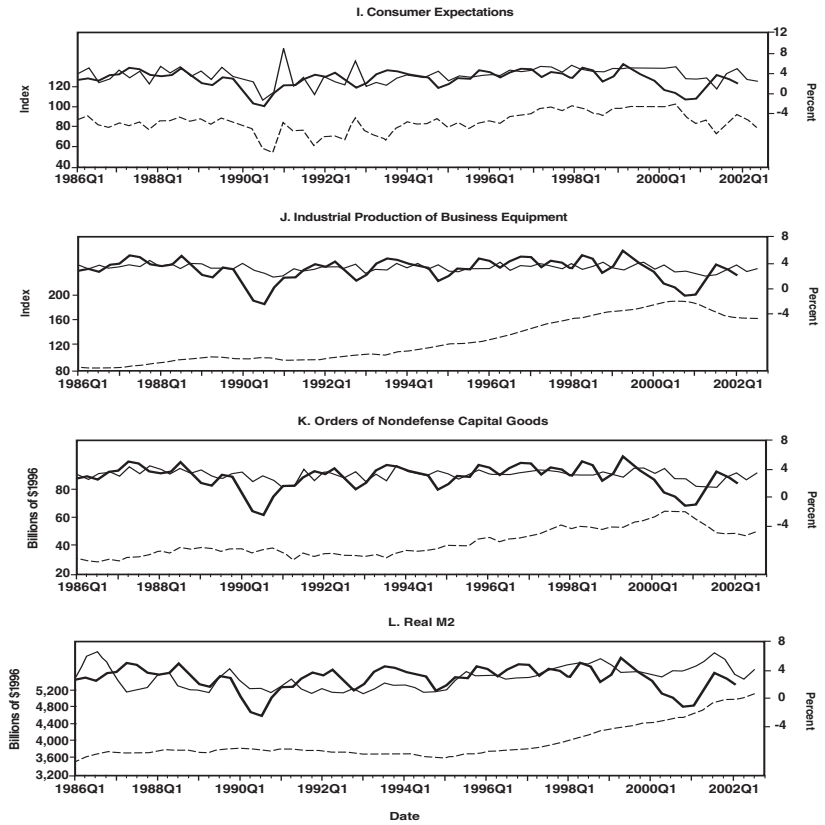
Figure 3 plots the time path of these twelve leading indicators from 1986Q1 through 2002Q3, along with actual two-quarter real GDP growth and its forecast based on that indicator. For each series in Figure 3, the solid lines are the actual two-quarter GDP growth (thick line) and its indicator-based forecast (thin line); the dates correspond to the date of the forecast (so the value plotted for the first quarter of 2001 is the forecasted and actual growth of GDP over the second and third quarters, at an annual rate). The dashed line is the historical values of the indicator itself (the value of the indicator plotted in the first quarter of 2001 is its actual value at that date). The scale for the solid lines is given on the right axis and the scale for the dashed line is given on the left axis.

Figure 3 Twelve Leading Indicators from 1986 to 2002, Two-Quarter Growth in Real GDP, and Its Leading-Indicator-Based Forecast



Inspection of Figure 3 reveals that some of these indicators moved in advance of the economic contraction, but others did not. The term spread provided a clear signal that the economy was slowing: the long government rate was less than the federal funds rate from June 2000 through March 2001. The decline in the stock market through the second half of 2000 also presaged further declines in the economy. New claims for unemployment insurance rose sharply over 2000, signaling a slowdown in economic activity. In contrast, other indicators, particularly series related to consumer spending, were strong throughout the first quarters of the recession. Housing starts fell sharply during the 1990 recession but remained strong through 2000. The consumer expectation series remained above 100 throughout 2000, reflecting overall positive consumer expectations. Although new capital goods orders dropped

Figure 3 Twelve Leading Indicators from 1986 to 2002, Two-Quarter Growth in Real GDP, and Its Leading-Indicator-Based Forecast



Notes: The solid lines are actual two-quarter GDP growth (thick line) and its indicator-based forecast (thin line), aligned so that the plotted date is the date of the forecast. The dashed line is the historical values of the indicator itself. The scale for the solid lines is given on the right axis and the scale for the dashed line is given on the left axis.

off sharply, that decline was contemporaneous with the decline in GDP, and in this sense new capital goods orders did not forecast the onset of the recession. The paper-bill spread provided no signal of the recession: although it moved up briefly in October 1998, October 1999, and June 2000, the spread was small and declining from August 2000 through the end of 2001, and the forecast of output growth based on the paper-bill spread remained steady and strong. In contrast, the junk bond spread rose sharply in 1998, leveled off, then rose

again in 2000. The junk bond spread correctly predicted a substantial slowing in the growth rate of output during 2001; however, it incorrectly predicted a slowdown during 1998. Finally, real M2 performed particularly poorly; the strong growth of the money supply before and during this recession led to M2-based output forecasts that were far too optimistic.

Quantitative Analysis of Forecast Errors

The graphical analysis shows that many of these indicators produced overly optimistic forecasts, which in turn led to large forecast errors. However, some indicators performed better than others. To assess forecast performance more precisely, we examine the mean squared forecast error over this episode of the different indicators relative to a benchmark autoregressive forecast. The mean squared forecast error is the most common way, but not the only way, to quantify forecasting performance, and we conclude this section with a brief discussion of the results if other approaches are used instead.

Relative Mean Squared Forecast Error

The relative mean squared forecast error (MSFE) compares the performance of a candidate forecast (forecast i) to a benchmark forecast; both forecasts are computed using the pseudo out-of-sample methodology. Specifically, let $\hat{Y}^h_{i,t+h|t}$ denote the pseudo out-of-sample forecast of Y^h_{t+h}, computed using data through time t, based on the ith individual indicator. Let $\hat{Y}^h_{0,t+h|t}$ denote the corresponding benchmark forecast made using the autoregression. Then the relative MSFE of the candidate forecast, relative to the benchmark forecast, is

$$\text{relative MSFE} = \frac{\sum_{t=T_1}^{T_2-h} (Y^h_{t+h} - \hat{Y}^h_{i,t+h|t})^2}{\sum_{t=T_1}^{T_2-h} (Y^h_{t+h} - \hat{Y}^h_{0,t+h|t})^2}, \tag{2}$$

where T_1 and $T_2 - h$ are, respectively, the first and last dates over which the pseudo out-of-sample forecast is computed. For this analysis, we set T_1 to 1999Q1 and T_2 to 2002Q3. If the relative MSFE of the candidate forecast is less than one, then the forecast based on that leading indicator outperformed the AR benchmark in the period just before and during the 2001 recession.

In principle, it would be desirable to report a standard error for the relative MSFE in addition to the relative MSFE itself. If the benchmark model is not nested in (that is, is not a special case of) the candidate model, then the standard error can be computed using the methods in West (1996). Clark and McCracken (2001) show how to test the hypothesis that the candidate model provides no improvement in the more complicated case that the candidate model nests the benchmark model. Unfortunately, neither situation applies

Table 4 Relative MSFEs of Combination Forecasts, 1999Q1–2002Q3

Combination Forecast Method	GDP		IP	
	$h = 2$	$h = 4$	$h = 2$	$h = 4$
Based on All Indicators				
Mean	0.95	0.94	0.95	0.95
Median	0.96	0.95	0.97	0.95
Inverse MSFE weights	0.97	0.98	0.95	0.96
Excluding Money				
Mean	0.94	0.91	0.91	0.92
Median	0.96	0.94	0.92	0.94
Inverse MSFE weights	0.96	0.95	0.93	0.94

Notes: Entries are the relative MSFEs of combination forecasts constructed using the full set of leading indicator forecasts in Table 3 (first three rows) and using the subset that excludes monetary aggregates (final three rows).

here because the lag length is chosen every quarter using the AIC; in some quarters the candidate model nests the benchmark, but in other quarters it does not. Because methods for this mixed case have yet to be worked out, the empirical results below report relative MSFEs but not standard errors.

Empirical Results

The relative MSFEs for thirty-seven leading indicators (including the twelve in Figure 3) are presented in the final four columns of Table 3 for two- and four-quarter-ahead forecasts of GDP growth and IP growth; the indicator and its transformation appear in the first two columns.

The mixed forecasting picture observed in Figure 3 is reflected in the MSFEs in Table 3. The relative MSFEs show that some predictors—the term spread, short-term interest rates, the junk bond spread, stock prices, and new claims for unemployment insurance—produced substantial improvements over the benchmark AR forecast. For example, the mean squared forecast error of the four-quarter-ahead forecast of GDP based on either measure of the term spread was one-third less than the AR benchmark. The two-quarter-ahead forecast of real GDP growth based on unemployment insurance claims had an MSFE 75 percent of the AR benchmark, another striking success.

In contrast, forecasts based on consumer expectations, housing starts, long-term interest rates, oil prices, or the growth of monetary aggregates all performed worse—in some cases, much worse—than the benchmark autoregression. Overall, the results from Table 3 reinforce the graphical analysis based on Figure 3 and provide an impression of inconsistency across indicators and, for a given indicator, inconsistency over time (e.g., the differing behavior of housing starts during the 1990 and 2001 recessions). This instability of

forecasts based on individual leading indicators is consistent with other recent econometric evidence on the instability of forecast relations in the United States and other developed economies; see, for example, the review of forecasts with asset prices in Stock and Watson (2001).

Results for Other Loss Functions

The mean squared forecast error is based on the most commonly used forecast loss function, quadratic loss. Quadratic loss implies a particular concern about large mistakes (a forecast error twice as large is treated as four times as "costly"). Although the theoretical literature abounds with other forecast loss functions, after quadratic loss the next most frequently used loss function in practice is mean absolute error loss, which in turn leads to considering the relative mean absolute forecast error (MAFE). The MAFE is defined in the same way as the MSFE in equation (2), except that the terms in the summation appear in absolute values rather than squared. The MAFE imposes less of a penalty for large forecast errors than does the MSFE.

We recomputed the results in Table 3 using the relative MAFE instead of the relative MSFE (to save space, the results are not tabulated here). The qualitative conclusions based on the relative MAFE are similar to those based on the relative MSFE. In particular, the predictors that improved substantially upon the AR as measured by the MSFE, such as the term spread and new claims for unemployment insurance, also did so as measured by the MAFE; similarly, those that fared substantially worse than the AR under the relative MSFE, such as consumer expectations and housing starts, also did so using the MAFE.

This analysis has focused on forecasts of growth rates. A different tack would be to consider forecasts of whether the economy will be in a recession, that is, predicted probabilities that the economy will be in a recession in the near future. This focus on recessions and expansions can be interpreted as adopting a different loss function, one in which the most important thing is to forecast the decree of the NBER Business Cycle Dating Committee. Because this episode has had only one turning point so far, the peak of March 2001, we think that more information about leading indicator forecasts during this period can be gleaned by studying quarterly growth rate forecasts than by focusing on binary recession event forecasts. Still, an analysis of recession event forecasts is complementary to our analysis, and recently Filardo (2002) looked at several probabilistic recession forecasting models. One of his findings is that the results of these models depend on whether final revisions or real-time data are used (the forecasts based on finally revised data are better). He also finds that a probit model based on the term spread, the paper-bill spread, and stock returns provided advance warning of the 2001 recession, a result consistent with the relatively good performance of the term spread and stock returns in Table 3.

3. COMBINATION FORECASTS

The SPF forecasts examined in Tables 1 and 2 are the average of the forecasts by the individual survey respondents. Such pooling of forecasts aggregates the different information and models used by participating forecasters, and studies show that pooled, or combination, forecasts regularly improve upon the constituent individual forecasts (see Clemen 1989; Diebold and Lopez 1996; and Newbold and Harvey 2002). Indeed, in their original work on leading indicators, Mitchell and Burns (1938) emphasized the importance of looking at many indicators, because each provides a different perspective on current and future economic activity.

In this section, we pursue this line of reasoning and examine the performance during the 2001 recession of combination forecasts that pool the forecasts based on the individual leading indicators examined in Section 3. The literature on forecast combination has proposed many statistical methods for combining forecasts; two important early contributions to this literature are Bates and Granger (1969) and Granger and Ramanathan (1984). Here we consider three simple methods for combining forecasts: the mean, the median, and an MSFE-weighted average based on recent performance.

The mean combination forecast is the sample average of the forecasts in the panel. The median modifies this by computing the median of the panel of forecasts instead of the mean, which has the potential advantage of reducing the influence of "crazy" forecasts, or outliers. This is the method that was used to produce the SPF combination forecasts in Table 1. The MSFE-weighted average forecast gives more weight to those forecasts that have been performing well in the recent past. Here we implement this combination forecast by computing the forecast error for each of the constituent forecasts over the period from 1982Q1 through the date that the forecast is made (thereby following the pseudo out-of-sample methodology), then estimating the current mean squared forecast error as the discounted sum of past squared forecast errors, with a quarterly discount factor of 0.95. The weight received by any individual forecast in the weighted average is inversely proportional to its discounted mean squared forecast error, so the leading indicators that have been performing best most recently receive the greatest weight.

The results are summarized in Table 4. The combination forecasts provide consistent modest improvements over the AR benchmark. During this episode, the simple mean performed better than either the median or inverse MSFE-weighted combination forecasts.

Because real money has been an unreliable leading indicator of output for many years in many developed economies (Stock and Watson 2001)—a characteristic that continued in the 2001 recession—it is also of interest to consider combination forecasts that exclude the monetary aggregates. Not surprisingly given the results in Table 3, the combination forecasts excluding money exhibit better performance than those that include the monetary aggregates.

Of course, the sample size is small and we should refrain from drawing strong conclusions from this one case study. Moreover, the improvements of the combination forecasts over the AR benchmark are less than the improvements shown by those individual indicators, such as new claims for unemployment insurance, that were, in retrospect, most successful during this episode. Still, the performance of the simple combination forecasts results is encouraging.

4. DISCUSSION AND CONCLUSIONS

Leo Tolstoy opened *Anna Karenina* by asserting, "Happy families are all alike; every unhappy family is unhappy in its own way." So too, it seems, with recessions. While the decline of the stock market gave some advance warning of the 2001 recession, it was not otherwise a reliable indicator during the 1980s and 1990s. Building permits and consumer confidence, which declined sharply preceding and during the 1990 recession, maintained strength well into the 2001 recession. While the term spread indicated an economic slowdown in 2001, it did not give an early signal in the 1990 recession. The varying performance of these indicators reflects the differences in the shocks and economic conditions prior to the 1990 and 2001 recessions.

In retrospect, the performance of the various individual indicators is generally consistent with the view that this recession was a joint consequence of a sharp decline of the stock market (perhaps nudged by some monetary tightening) and an associated pronounced decline in business investment, especially in information technology. These shocks affected manufacturing and production but diffused only slowly to general employment, incomes, and consumption. But without knowing these shocks in advance, it is unclear how a forecaster would have decided in 1999 which of the many promising leading indicators would perform well over the next few years and which would not.

The failure of individual indicators to perform consistently from one recession to the next, while frustrating, should not be surprising. After all, the U.S. economy has undergone important changes during the past three decades, including an expansion of international trade, the development of financial markets and the concomitant relaxing of liquidity constraints facing consumers, and dramatic increases in the use of information technology in manufacturing and inventory management. Moreover, the conduct of monetary policy arguably has shifted from being reactionary, using recessions to quell inflation, to more proactive, with the Fed acting as if it is targeting inflation (see Goodfriend 2002). As we discuss elsewhere (Stock and Watson 2001, 2003a), these and other macroeconomic changes could change the relation between financial leading indicators and economic activity and, to varying degrees, could contribute to the reduction in volatility of GDP that the United States (and other countries) have enjoyed since the mid-1980s.

Our conclusion—that every decline in economic activity declines in its own way—is not new. Indeed, one of the reasons that Mitchell and Burns (1938) suggested looking at many indicators was that each measured a different feature of economic activity, which in turn can play different roles in different recessions. In light of the variable performance of individual indicators and the evident difficulty professional forecasters had during this episode, the results for the combination forecasts are encouraging and suggest that, taken together, leading economic indicators did provide some warning of the economic difficulties of 2001.

REFERENCES

Bates, J. M., and Clive W. J. Granger. 1969. "The Combination of Forecasts." *Operations Research Quarterly* 20: 451–68.

Blanchard, Olivier J. 1993. "Consumption and the Recession of 1990–1991." *American Economic Review Papers and Proceedings* 83(2): 270–75.

Clark, Todd E., and Michael W. McCracken. 2001. "Tests of Equal Forecast Accuracy and Encompassing for Nested Models." *Journal of Econometrics* 105 (November): 85–100.

Clemen, Robert T. 1989. "Combining Forecasts: A Review and Annotated Bibliography." *International Journal of Forecasting* 5(4): 559–83.

Diebold, Francis X., and J. A. Lopez. 1996. "Forecast Evaluation and Combination." In *Handbook of Statistics*, edited by G. S. Maddal and C. R. Rao. Amsterdam: North-Holland.

Estrella, Arturo, and Gikas Hardouvelis. 1991. "The Term Structure as a Predictor of Real Economic Activity." *Journal of Finance* 46 (June): 555–76.

Filardo, Andrew J. 2002. "The 2001 U.S. Recession: What Did Recession Prediction Models Tell Us?" Manuscript, Bank for International Settlements.

Friedman, Benjamin M., and Kenneth N. Kuttner. 1992. "Money, Income, Prices and Interest Rates." *American Economic Review* 82 (June): 472–92.

Gertler, Mark, and Cara S. Lown. 2000. "The Information in the High Yield Bond Spread for the Business Cycle: Evidence and Some Implications." NBER Working Paper 7549 (February).

Goodfriend, Marvin. 2002. "The Phases of U.S. Monetary Policy: 1987 to 2001." Federal Reserve Bank of Richmond *Economic Quarterly* 88 (Fall): 1–17.

Granger, Clive W. J., and Ramu Ramanathan. 1984. "Improved Methods of Combining Forecasting." *Journal of Forecasting* 3: 197–204.

Hall, Robert E. 2002. "Dating Business Cycles: A Perspective." Manuscript, Stanford University.

Harvey, Campbell R. 1988. "The Real Term Structure and Consumption Growth." *Journal of Financial Economics* 22 (December): 305–33.

_____. 1989. "Forecasts of Economic Growth from the Bond and Stock Markets." *Financial Analysts Journal* 45 (September/October): 38–45.

Hamilton, James D. 1983. "Oil and the Macroeconomy Since World War II." *Journal of Political Economy* 91 (April): 228–48.

Merrill Lynch. www.mlindex.ml.com.

Mitchell, Wesley C., and Arthur F. Burns. 1938. *Statistical Indicators of Cyclical Revivals*, NBER Bulletin 69. Reprinted as Chapter 6 of *Business Cycle Indicators,* edited by G. H. Moore. Princeton, N.J.: Princeton University Press, 1961.

Newbold, Paul, and David I. Harvey. 2002. "Forecast Combination and Encompassing." In *A Companion to Economic Forecasting,* edited by M. P. Clements and D. F. Hendry. Oxford: Blackwell Press.

Stock, James H., and Mark W. Watson. 1989. "New Indexes of Coincident and Leading Economic Indicators." In *NBER Macroeconomics Annual 1989*, edited by O. J. Blanchard and S. Fischer. Cambridge: MIT Press, 352–94.

_____. 2001. "Forecasting Output and Inflation: The Role of Asset Prices." NBER Working Paper 8180 (March).

_____. 2003a. "Has the Business Cycle Changed and Why?" In *NBER Macroeconomics Annual 2002*, edited by Mark Gertler and Kenneth Rogoff. Cambridge: MIT Press.

_____. 2003b. *Introduction to Econometrics*. Boston: Addison-Wesley.

West, Kenneth D. 1996. "Asymptotic Inference About Predictive Ability." *Econometrica* 64 (September): 1067–84.