Federal Reserve Bank of New York

March 2001 Volume 7 Number 1

Economic Policy Review

Note that we have changed our web site address. Please update your bookmarks.

www.newyorkfed.org/rmaghome

The *Economic Policy Review* is published by the Research and Market Analysis Group of the Federal Reserve Bank of New York. Articles undergo a comprehensive refereeing process prior to their acceptance in the *Review*. The views expressed in the articles are those of the individual authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

Federal Reserve Bank of New York Economic Policy Review

March 2001
Volume 7 Number 1

Contents

Christine M. Cumming and Beverly J. Hirtle

The Challenges of Risk Management in Diversified Financial Companies

- Although the benefits of a consolidated, or firmwide, system of risk management are widely recognized, financial firms have traditionally taken a more segmented approach to risk measurement and control.
- The cost of integrating information across business lines and the existence of regulatory barriers to moving capital and liquidity within a financial organization appear to have discouraged firms from adopting consolidated risk management.
- In addition, there are substantial conceptual and technical challenges to be overcome in developing risk management systems that can assess and quantify different types of risk across a wide range of business activities.
- However, recent advances in information technology, changes in regulation, and breakthroughs in risk management methodology suggest that the barriers to consolidated risk management will fall during the coming months and years.

In recent years, financial institutions and their supervisors have placed increased emphasis on the importance of consolidated risk management. Consolidated risk management—sometimes also called integrated or enterprisewide risk management—can have many specific meanings, but in general it refers to a coordinated process for measuring and managing risk on a firmwide basis. Interest in consolidated risk management has arisen for a variety of reasons. Advances in information technology and financial engineering have made it possible to quantify risks more precisely. The wave of mergers—both in the United States and overseas—has resulted in significant consolidation in the financial services industry as well as in larger, more complex financial institutions. The recently enacted Gramm-Leach-Bliley Act seems likely to heighten interest in consolidated risk management, as the legislation opens the door to combinations of financial activities that had previously been prohibited.

This article examines the economic rationale for managing risk on a firmwide, consolidated basis. Our goal is to lay out some of the key issues that supervisors and risk management practitioners have confronted in assessing and developing consolidated risk management systems. In doing so, we hope to clarify for a wider audience why the ideal of consolidated risk management—which may seem uncontroversial or even obvious—involves significant conceptual and practical issues. We also hope to suggest areas where research by practitioners and academics could help resolve some of these issues.

Christine M. Cumming is an executive vice president and the director of research and Beverly J. Hirtle is a vice president at the Federal Reserve Bank of New York.

The authors would like to thank Gerald Hanweck, Darryll Hendricks, Chris McCurdy, Brian Peters, Philip Strahan, Stefan Walter, Lawrence White, and two anonymous referees for many helpful comments. The views expressed are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

The approach we take is to review the arguments made by supervisors and the financial industry in favor of consolidated risk management. While both parties agree on the importance of this type of risk management, this support seems to be motivated by quite different concerns. Supervisors appear to support it out of a safety-and-soundness concern that significant risks could be overlooked or underestimated in the absence of firmwide risk assessment.[1] In contrast, financial institutions appear willing to undertake significant efforts to develop consolidated risk management systems because they believe that those systems will help them assess the risk and return of different business lines and thus allow them to make more informed decisions about where to invest scarce resources to maximize profits.[2] While these two views may reflect quite different underlying motivations for supporting consolidated risk management, we argue below that they result in a common emphasis on the importance of accurate assessments of risk.

Our goal is to lay out some of the key issues that supervisors and risk management practitioners have confronted in assessing and developing consolidated risk management systems.

Although both supervisors and financial institutions support the concept of consolidated risk management, few if any financial firms have fully developed systems in place today. The absence thus far of fully implemented consolidated risk management systems suggests that there are significant costs or obstacles that have historically led firms to manage risk in a more segmented fashion. We argue that both information costs and regulatory costs play an important role here by affecting the trade-off between the value derived from consolidated risk management and the expense of constructing these complex risk management systems. In addition, there are substantial technical hurdles involved in developing risk management systems that span a wide range of businesses and types of risk. Both of these factors are evolving in ways that suggest that the barriers to consolidated risk management are increasingly likely to fall over the coming months and years.

The remainder of this article is organized as follows. In the next section, we describe the concept of consolidated risk management in greater detail and provide a more in-depth discussion of the views of supervisors and the financial industry about this process. We then offer a critical analysis of these views, using a simple portfolio model to help illustrate the economic rationale behind consolidated risk management. Next, we discuss the constraints that have slowed many financial institutions in their implementation of consolidated risk management systems. We conclude with a discussion of the major technical challenges and research questions that will need to be addressed as an increasing number of financial firms implement firmwide risk management systems.

Consolidated Risk Management: Definitions and Motivations

At a very basic level, consolidated risk management entails a coordinated process of measuring and managing risk on a firmwide basis. This process has two distinct, although related, dimensions: coordinated risk assessment and management across the different types of risk facing the firm (market risk, credit risk, liquidity risk, operational risk), and integrated risk evaluation across the firm's various geographic locations, legal entities, and business lines. In theory, both dimensions must be addressed to produce a consolidated, firmwide assessment of risk. In practice, few financial firms currently have in place a consolidated risk management system that fully incorporates both dimensions, although many large institutions—both in the United States and overseas—appear to be devoting significant resources to developing such systems (Joint Forum 1999a).[3]

The absence thus far of fully implemented consolidated risk management systems suggests that there are significant costs or obstacles that have historically led firms to manage risk in a more segmented fashion.

To understand consolidated risk management, it is important to recognize the distinction between risk measurement and risk management. *Risk measurement* entails the quantification of risk exposures. This quantification may take a variety of forms—value-at-risk, earnings-at-risk, stress scenario analyses, duration gaps—depending on the type of risk being measured and the degree of sophistication of the estimates. *Risk management*, in contrast, refers to the overall process that a financial institution follows to define a business strategy, to identify the risks to which it is exposed, to quantify those risks, and to understand and control the nature of the risks it faces. Risk management is a series of business decisions, accompanied by a set of checks and balances—risk limits, independent risk management functions, risk reporting, review and oversight by senior management and the board of directors—in which risk measurement plays an important, although not all-encompassing, role. Thus, consolidated risk management involves not only an attempt to quantify risk across a diversified firm, but also a much broader process of business decision making and of support to management in order to make informed decisions about the extent of risk taken both by individual business lines and by the firm as a whole.

Recent trends in the financial services industry have increased the challenges associated with this process. To begin, financial institutions increasingly have the opportunity to become involved in a diverse range of financial activities. In the United States, bank holding companies have been able to combine traditional banking and securities activities since the late 1980s, when the Federal Reserve permitted the creation of "Section 20" securities subsidiaries. The Gramm-Leach-Bliley Act will now enable affiliations involving banking, securities, and insurance underwriting in so-called financial holding companies (FHCs). Such combinations of diverse financial activities present significant challenges to consolidated risk management systems, as greater diversity often means that the system must encompass a wider range of risk types.[4]

Consolidation in the financial services industry has produced institutions with operations spanning large geographic areas, both domestically and internationally. Such wide geographic dispersion, especially across time zones, can make it difficult for a firm's management to keep track of the activities across all of its operating centers. Financial institutions have responded to this situation by increasing the resources devoted to information systems designed to track and monitor exposures worldwide. Indeed, the development of coordinated information systems is one of the most important steps in consolidated risk management.

The supervisory community has advocated that financial institutions adopt consolidated risk management procedures in the guidance it has published in the 1990s, especially guidance for banking companies. In the United States, these efforts began in 1993 with guidelines for supervisors evaluating risk management in derivatives and trading activities, and have continued to date, most recently with a 1999 Federal Reserve paper containing broad conceptual guidelines for evaluating capital adequacy in light of the full range of risks facing the bank or bank holding company.[5] Internationally, the Basel Committee on Banking Supervision extended the framework for describing the risk management process to encompass the role of business strategy and the activities of business line decision makers.[6] The Committee also set out an approach to the supervisory review of a bank's internal assessment of capital adequacy in light of a firm's overall risks as the second pillar of the proposed new capital adequacy framework (Basel Committee on Banking Supervision 1999b).

Recently, an international forum of banking, securities, and insurance supervisors issued a report containing principles that supervisors should follow to ensure that financial conglomerates are adequately identifying and managing risk. The report's lead recommendation is that "supervisors should take steps . . . to provide that conglomerates have adequate risk management processes in place to manage group-wide risk concentrations" (Joint Forum 1999a).

The rationale offered by supervisors for the importance of consolidated risk management seems to be a concern that, in the absence of a firmwide assessment, significant risks could be overlooked or underestimated. The Joint Forum report, for instance, argues that "the additive nature of concentrations and the risk of transmission of material problems within a conglomerate point to the value of both conglomerate management and supervisors conducting a group-wide assessment of potential concentrations" (Joint Forum 1999a). The underlying concern is that such underestimated or overlooked risks receive insufficient management attention and have the potential to produce unexpectedly large losses that could threaten the firm's financial health.

Consolidated risk management involves not only an attempt to quantify risk across a diversified firm, but also a much broader process of business decision making and of support to management in order to make informed decisions about the extent of risk taken both by individual business lines and by the firm as a whole.

Financial market practitioners also cite the interdependent nature of risks within an organization as a motivation to develop consolidated risk management systems. For instance, echoing sentiments in the supervisors' Joint Forum report, Lam (1999) argues that "managing risk by silos simply doesn't work, because the risks are highly interdependent and cannot be segmented and managed solely by independent units" in the firm. Similarly, a senior executive at a major U.S. bank asserts that "the careful identification and analysis of risk are, however, only useful insofar as they lead to a capital allocation system that recognizes different degrees of risk and includes all elements of risk" (Labrecque 1998).

In contrast to the supervisors, however, the primary implication that Lam and others draw from this finding concerns the role that consolidated risk management systems can play in helping firms to make better-informed decisions about how to invest scarce capital and human resources. For instance, Mudge (2000) stresses that a consistent framework for evaluating firmwide risk and return across diverse financial activities is a key to evaluating the benefits of potential mergers among banking and insurance firms. Similarly, Lam (1999) argues that consolidated risk management systems can help firms understand the risk/return trade-offs among different business lines, customers, and potential acquisitions. Furthermore, consolidated risk management may allow a firm to recognize "natural hedges"—when one entity within the firm has positions or is engaged in activities that hedge the positions or activities of another part of the firm—that may become apparent only when risk is examined from the perspective of the consolidated institution. Firms that fail to recognize the diversification effects of such natural hedges may waste resources on redundant hedging by individual units within the organization.

While both supervisors and financial institutions agree on the importance of consolidated risk management and point to the same driving factors, their conclusions about the role that these systems can play emphasize quite different concerns.

Thus, while both supervisors and financial institutions agree on the importance of consolidated risk management and point to the same driving factors, their conclusions about the role that these systems can play emphasize quite different concerns. At one level, this difference is not surprising, given the different objectives of supervisors and financial institutions (safety and soundness, on the one hand, and profit maximization, on the other). On another level, these concerns are not necessarily mutually exclusive. Indeed, in the next section, we argue that supervisors' emphasis on underestimation of firmwide risk and financial institutions' emphasis on enhanced understanding of the risk/return trade-off among different activities reflect a common emphasis on the importance of accurate assessments of risk.

Understanding the Role of Consolidated Risk Management

The discussion above reflects a well-established belief on the part of financial institutions and supervisors in the importance of consolidated risk management. But what economic fundamentals underlie this belief? In this section, we assess the views of supervisors and financial institutions and try to place them in a common framework. We do not attempt to address the question of why firms choose to manage risk at all.[7] Instead, we try to understand why it matters whether risk is managed on a consolidated basis or at the level of individual businesses or risks within a firm.

The Supervisors' View: Spillover Effects

We first consider the view expressed by supervisors in the Joint Forum paper (1999a), namely, that in the absence of consolidated risk management, significant risks could be overlooked or underestimated. To gain some insight into this view, it is helpful to consider a simple portfolio approach to assessing the risk of a diversified financial firm. This approach helps illustrate how the perception of the overall risk facing the firm would differ if institutions managed their risk in an integrated way instead of by individual businesses or legal entities within the larger organization.

To begin, suppose that a financial firm has two business lines, each of which earns profits that vary uncertainly over time. Application of standard portfolio theory suggests that the risk of the overall firm will depend on the variation in each unit's profits and the extent to which variation in these profits is correlated between the two units. In particular, the risk facing the consolidated firm will be less than or equal to the sum of the risks facing the individual business units within the firm whenever this correlation is less than perfect. In this situation, the profit variation in one unit diversifies the risk of the other.[8]

The importance of this observation for our purposes is that it suggests that establishing risk monitoring and control (such as limits) at the business level and then summing up across business lines would be a *conservative* approach to managing and assessing the overall risk facing the firm, since it ignores any potential diversification effects across business lines. This conclusion stands in marked contrast to the arguments advanced by supervisors in favor of consolidated risk management. How can we reconcile these two outcomes?

The answer, of course, is that the simple portfolio example misses some important "real world" aspects of financial risk and risk management. Perhaps the most significant of these is the assumption that the risks facing each business unit are fixed and known. In fact, these risks are functions of many factors that can vary significantly over time. In particular, the simple portfolio example assumes that the risk profile of one business line can be measured without regard to the risks undertaken by the other. This assumption is not a statement about the degree of correlation between the risks faced by the two business units, but rather the idea that the underlying volatility of one business line's profitability may be affected by the actions of another business line.

Spillover effects can be enhanced during times of crisis or severe market disruption. A firm that manages risk on a unit-by-unit basis may have to spend valuable time simply determining what its aggregate position is in the affected markets, rather than being able to react to quickly developing market conditions.

An example of this relationship might be when two or more geographic centers within a global financial firm have similar positions that they have each hedged in a particular security or market. In the absence of a consolidated risk management system, the various units could be unaware of the positions that other units within the firm have taken.[9] Each unit assumes that its position is small enough that it would be able to roll over its hedges or otherwise take steps to reduce its risk even in the event of market stress. However, when the various business units try to take these steps simultaneously, their combined activity reinforces the liquidity problems facing the market, resulting in sharp, adverse moves in the market prices of the hedging and/or underlying instruments. Thus, losses at individual units exceed the risk assumptions made in each unit's individual risk management plans and the aggregate position of the firm is therefore riskier than the sum of the *assumed* individual risks of the business units. In essence, the firm faces the "portfolio insurance" problem in that the actions of one unit affect the risks facing another.[10]

These spillover effects can be enhanced during times of crisis or severe market disruption. A firm that manages risk on a unit-by-unit basis may have to spend valuable time simply determining what its aggregate position is in the affected markets, rather than being able to react to quickly developing market conditions. Since nimbleness in responding to problems can affect outcomes favorably, such firms may be at a disadvantage compared with smaller firms (for instance, compared with a series of smaller firms that are comparable in the aggregate to the diversified financial institution) and compared with large firms with consolidated risk management systems. Such a situation is an example of how the structure of the risk management system—as distinct from any ex ante risk-mitigating actions taken by the firm's risk managers—may affect the aggregate risk facing the firm. Nimbleness is especially important if market disruption spreads rapidly from market to market in a hard-to-anticipate pattern, as it did in 1997-98.

In fact, the financial crisis in the fall of 1998 provides some interesting insights into the importance of consolidated risk management and measurement systems when there are linkages across markets. International bank supervisors conducted a study of the performance during the market upheaval of banks' risk management systems and the value-at-risk models used to calculate market risk capital requirements (Basel Committee on Banking Supervision 1999c). The study examined information on the stress testing done by large banks in several G-10 countries and found that ex ante stress test results provided a better picture of actual outcomes during the third quarter of 1998, when those tests were based on actual historical experience or hypothetical scenarios that incorporated simultaneous movements in a range of rates and prices, rather than on large movements in a single market risk factor. Thus, firms whose stress testing and risk management systems recognized potential linkages across markets had more realistic estimates of the way events in the fall of 1998 were likely to affect their firms.

Another way in which spillover effects can result in aggregate risk exceeding the sum of the individual risks of business units within the firm concerns what might be called reputational or contagion risk. As discussed in the Joint Forum report (1999a), this is the idea that problems in one part of a diversified firm may affect confidence in other parts of the firm. The situation that the Joint Forum paper appears to have in mind is one in which such problems cause acute, near-term funding or liquidity problems across the firm, due to questions about whether the losses in the troubled business unit are evidence of as-yet-unrevealed losses in other business lines.[11]

Aside from such near-term concerns, spillover effects can also have a longer run dimension. For example, innovative businesses or those involving massive technology investments can engender what some analysts call "strategic risk." Failure in such ventures may be highly visible and thus likely to have spillover effects on other businesses through the cost of capital, the cost of funding, and revenue effects through the loss of customer approval. Thus, other business lines associated with the troubled entity may see their franchise value erode as a result of difficulties in an affiliated unit. Such strategic risk may be particularly important for institutions for which customer trust is a key competitive advantage. Adverse publicity, legal judgments against the firm, evidence of fraud or internal theft, or high-profile failed business ventures may erode customer confidence in an institution. In the extreme, such concerns may reach the point where the affected firm is no longer viable as an ongoing concern, even though it may technically be solvent.[12]

Certain important risks may be very difficult, if not impossible, to incorporate into risk management systems that focus on individual business units or types of risk alone within a diversified firm.

This discussion of spillover effects suggests that supervisors' concerns that disaggregated risk management systems understate the risks facing diversified financial institutions may not be without foundation. Certain important risks may be very difficult, if not impossible, to incorporate into risk management systems that focus on individual business units or types of risk alone within a diversified firm. Consolidated risk management systems therefore may be necessary to obtain an accurate picture of the risks facing a firm and to have in place the procedures needed to manage those risks, both on a day-to-day basis and in stress situations. In this light, supervisors' concerns can be seen not so much as a desire for firms to have risk management systems that are conservative, but instead for firms to have risk management systems that are accurate.

Consolidated Risk Management and the Theory of the Firm

Concerns about understating firmwide risk exposures notwithstanding, disaggregated risk management systems may also miss instances in which the risks from different units within a diversified firm offset one another. The consolidated firm would appear to have incentives to manage its risk on an aggregate basis whenever these diversification benefits are non-negligible. At its heart, this is the logic that Lam and others in the financial services industry have applied in support of consolidated risk management: the idea that a diversified financial firm should be viewed as a "portfolio" comprising its different units and business lines.

This view is closely related to the broader question of how firms decide which activities are coordinated within the firm and which activities are coordinated through markets. This question has long interested economists, and we can draw on the insights of this "theory of the firm" literature to enhance our understanding of the role of consolidated risk management. Coase (1937) first noted that the efficiency of markets might be expected to lead firms to rely on markets and contracts with third parties to conduct their activities, but that in fact many decisions are made, coordinated, and executed by internal mechanisms such as reporting hierarchies, production organization, and compensation plans. Coase's insight was that a firm carries out inside the firm those activities that it can manage internally at a cost lower than the information and transaction costs involved in purchasing corresponding services or goods outside the firm.

Since the mid-1970s, economists have further developed and extended the Coase analysis by elaborating more fully on the roles of contracting for goods and services and the ownership of assets in determining what is coordinated within the firm and what is coordinated by markets. Grossman and Hart (1986) noted that the combination of uncertainty and complexity makes contracting with inside or outside parties difficult. In the presence of less than fully specified contracts, ownership and control of assets is synonymous with ownership of the rights not otherwise covered by contract. Thus, the ease or difficulty of contracting plays a major role in determining what occurs inside the firm. Ownership demarcates the boundary of the firm's internal activities, which often involve the "noncontractible" aspects of the firm's activities. In the Grossman and Hart analysis, bringing activities under common ownership (integration) makes economic sense whenever efficiency gains from improved information and coordination within the firm exceed the efficiency losses resulting from the reduced entrepreneurial incentive of the manager who is no longer an owner.

The basic implication of this literature is that activities will be performed inside the firm when the complexity or costs of performing them outside the firm are high. For a diversified financial firm, these insights can be applied to interactions between the various units within the firm. In this setting, we can think of activities conducted by a corporate parent on a firmwide basis as coordination "inside" the firm, while activities conducted independently by separate units of the firm are analogous to the "market" activities discussed in Coase and in Grossman and Hart. Following this logic, risk management and other corporate control activities will be conducted on a consolidated basis when it is too difficult or costly for the individual business units to contract among themselves.

The type of spillover effects and interrelated risks discussed above arguably create just such a situation. When the actions of one business unit in a diversified firm potentially affect the risks faced by others, the contracting problem—in this case, what risk exposures may be undertaken by the various business units within the firm—becomes very complex to solve on a bilateral basis. In such circumstances, the incentives to create a centrally run, consolidated risk management system may be strong.

Fungibility of Financial Resources

That consolidated risk management allows the firm to allocate capital efficiently further reinforces the interdependence between a firm's business units. The fungibility of capital within the firm—what some have called a firm's internal capital market—means that the risks undertaken by one unit can affect the resources available to another through the workings of the internal capital market. In considering risk in relation to the capital resources available to back that risk, then, an additional dimension is that those resources may also be called into play to back the activities of other units within the firm.[13]

The financial institution's internal capital market is itself an example of coordination within the firm potentially being more efficient than external markets. Gertner, Scharfstein, and Stein (1994) attribute the efficiency of internal capital markets to the strong incentive that owners have to monitor capital use relative to debtholders, especially if many aspects of the firm's capital use are not limited by the debtholders' contract. In addition, capital allocated to an unsuccessful project can be shifted to another use within the firm at less cost than would be involved in liquidating the assets of the project in the market, if capital and resources in one use are close enough substitutes for those in other activities. As discussed earlier, these benefits are offset by a reduction in incentives to managers who no longer act like owners.

Froot and Stein (1998) offer a model of capital allocation and capital structure for financial firms that develops the relationship between risk management and capital allocation formally.[14] In their model, financial institutions fully hedge risks for which liquid markets are available. Financial institutions have incentives to engage in risk management whenever they face risks that cannot be traded in liquid markets because they need to hold capital against the nontradable positions according to the amount of risk in the portfolio.[15] The desirability of any given investment depends on the extent to which its nontradable risk is correlated with the nontradable risks of the firm's other portfolio positions. Drawing this point to its logical conclusion, Froot and Stein argue that "this line of reasoning suggests that the right question is not whether or not the bank should centralize its decisionmaking, but rather how often headquarters should gather information and use this pooled information to help guide investment decisions."

That consolidated risk management allows the firm to allocate capital efficiently further reinforces the interdependence between a firm's business units.

The firm's liquidity resources (assets that can be liquidated as well as funding sources that can be tapped) can be viewed as fungible across the firm in much the same way that capital is fungible (in the absence of regulatory or other constraints). For this reason, liquidity resources virtually always are coordinated centrally for the firm as a whole (Basel Committee on Banking Supervision 2000a). These resources are available to provide cash needed to meet obligations, especially in contingency situations such as market distress.

This interdependency suggests that consolidated risk management systems should take liquidity considerations into account. Liquidity risk assessment requires knowledge of the size and nature of the firm's risk positions, while the firm's liquidity risk position should influence the amount and type of risk that business managers choose to take. One approach to recognizing this connection is to extend the concept of capital adequacy to encompass the ability to liquidate assets or easily fund them, as is intended by the Securities and Exchange Commission's capital rule for registered broker-dealers. Alternatively, an integrated risk assessment approach could consider liquidity risk along with market, credit, and other risks in scenario analyses intended to test the impact of the scenario on capital adequacy (and ultimately solvency) and liquidity, in a test of dual constraints.

Finally, the risks introduced by leverage reinforce the need to evaluate risk on a firmwide basis. Most financial firms are leveraged, and over the course of the 1990s analysts in financial institutions and their supervisors have recognized that many methods can be used to increase leverage in addition to increasing balance sheet debt to equity, such as taking positions through the use of derivatives and imbedded optionality. Since leverage increases the risk supported by capital, a sophisticated risk assessment should incorporate the combined effects of all sources of market and credit risks, of liquidity risk, and of leverage on capital. This point was made by the Counterparty Risk Management Policy Group (1999) in its private sector report on lessons learned from the 1998 de facto failure of Long-Term Capital Management, a large hedge fund. The report suggests several measures that can be used to conduct a risk and capital or liquidity adequacy analysis.

Debtholders and Other Creditors

Financial institutions may have additional incentives to engage in consolidated risk management because of the concerns of debtholders and other creditors.[16] In agreeing to extend credit, these parties must take into account the moral hazard incentive that the firm has to increase its risk exposure—to the benefit of the firm's shareholders and the detriment of its creditors—once the credit has been extended. This situation is particularly acute for financial firms, which can change their risk profiles relatively rapidly using derivatives and other liquid financial instruments. In the face of this uncertainty, creditors may charge higher rates or offer less favorable nonprice terms (for instance, shorter maturity or higher collateral) than they would if this incentive could be addressed.

Consolidated risk management systems provide a way for financial institutions to make a credible commitment against such behavior. In particular, these systems facilitate better disclosure by providing a consistent and comprehensive assessment of the firm's true risk exposure that can be used by creditors to monitor the institution's activities. In the absence of such systems, it can be significantly more difficult for analysts to draw an accurate picture of the firm's overall risk exposure, even if the individual units within the firm make extensive disclosures of their risk profiles. Furthermore, the centralized and independent risk management units that nearly always are a key feature of consolidated risk management systems provide an internal check against any incentives for individual units or employees within the firm to hide risk exposures from senior management. Finally, the enhanced disclosure made possible by consolidated risk management systems may mitigate some of the spillover effects described above by providing meaningful information about the true extent and nature of linkages between various businesses within the consolidated firm.[17] Thus, these systems can provide an important tool for management to address the moral hazard concerns of creditors and to obtain better borrowing terms as a result.

Spillover effects, the fungibility of resources, and the concerns of debtholders and creditors suggest that firms have strong incentives to measure risks well, to take advantage of diversification benefits, and to manage capital and liquidity efficiently. In the next section, we examine why firms have not been faster to adopt consolidated risk management to take advantage of even small diversification effects and why both industry and supervisory efforts have been necessary to encourage its use.

Consolidated risk management systems . . . facilitate better disclosure by providing a consistent and comprehensive assessment of the firm's true risk exposure that can be used by creditors to monitor the institution's activities.

Obstacles to Creating Consolidated Risk Management Systems

That firms have not immediately adopted consolidated risk management systems suggests that there are significant costs or obstacles that historically have led firms to manage risk in a more segmented fashion. While the firm can invest in two business activities, as discussed above, it finds the two activities to be in some sense segregated, so that taking advantage of diversification effects engenders costs. The segregation can be geographical (such as New York versus London) or conceptual (for example, loans versus over-the-counter options).

Information Costs

Segregation creates two kinds of costs. The first is information costs—the costs of integrating and analyzing information from the two business lines. Those costs involve both the resources involved in transmitting, recording, and processing the information and the amount of decay in the time value of the information, reflecting the lags in assembling and verifying information. At any given moment, there may be competing information technologies with similar scale effects, but a different mix of costs in terms of monetary outlays and time to assemble information (for instance, a highly automated process versus a manual one).

Information costs are shaped largely by technology. Information systems tend to have substantial fixed costs that usually increase with the size of the information system, but low marginal costs until the particular system approaches capacity. To reflect that, we consider the total information cost function to be a step function increasing discretely as the scale of the business increases. For a given volume of information, then, the value of recognizing the impact of diversification—which is a function of the amount of diversification inherent in the firm's activities—needs to exceed the information costs for the scale of the firm's business in order for the firm to invest in the information infrastructure. In essence, the firm maximizes its expected profits subject to a capital constraint by choosing the business mix, the scale of business, and the information technology (or none) to manage risk.[18]

Improvements in technology reduce fixed information costs, make it possible for firms to take greater advantage of diversification benefits, and increase the scale on which certain businesses can be conducted.

Information costs will tend to limit the size of the business for a given level of capital. If the firm finds the cost of information high relative to the diversification benefit, the firm will manage each business separately, and in doing so, it will assign relatively high amounts of capital to each business line as if there was no diversification benefit. As a result, the scale of the firm's overall business will be lower than it would be when diversification effects can be realized.[19]

Improvements in technology reduce fixed information costs, make it possible for firms to take greater advantage of diversification benefits, and increase the scale on which certain businesses can be conducted. For example, improvements in information technology permit banks and securities firms to manage single "global books," in contrast to the regional approach used to manage most international businesses in the 1970s and 1980s.

Finally, the value of information has risen as the pace of developments has picked up and the complexity of financial relationships among markets and counterparties has increased. If we interpret the increased speed of events as an increase in the variability of the risks and correlations associated with a financial firm's different business lines, then, *ceteris paribus*, firms would tend to set necessarily more conservative limits on their activities—perhaps in line with the maximum possible values of the risk exposures of their various units.[20] Since these maxima would rarely be observed together in practice, there would appear to be substantial opportunities for gains from identifying and responding to changes in the diversification benefit. But greater volatility in the underlying risk relationships also changes the set of relevant information technologies, since at any scale of activity most "low-tech," time-intensive techniques become unacceptably costly, reflecting the rapid decay in the value of information. Thus, in a more volatile environment, we might expect the ability to design and implement effective technology-intensive risk management information systems to represent a significant dimension of competitiveness for financial institutions seeking to operate in a large number of markets.[21]

Regulatory Costs

Regulatory barriers to moving capital and liquidity within a financial organization impose another cost that inhibits the use of consolidated risk management. These barriers can take the form of business line capital and liquidity requirements set by regulators, prohibitions or limits on capital and funds that can be transferred from one business line to another, or the necessity of seeking prior approval or giving prior notice to move funds between business lines. Most commonly, business lines segregated from one another by such regulatory requirements are in different locations or different legal entities, subjecting the two business lines to different regulations. However, similar types of costs can be imposed by rating agencies, creditors, or even investors when the requirements or expectations they set differ across individual entities.

As with information costs, we can consider the regulatory costs to reflect both monetary outlays to manage or circumvent regulatory barriers and the waiting period or decay in profit opportunities in the time needed to comply with or overcome regulatory costs. While in some cases regulatory requirements can make it virtually impossible to move capital or liquidity from one business line to another in the short run, in many cases regulatory requirements can be satisfied at some cost. The cost of managing and circumventing regulatory requirements appears to have dropped substantially through the use of derivatives, securitization techniques, and other financial engineering. Indeed, a recurring pattern in financial regulation is the erosion of regulatory requirements through financial innovation and regulatory arbitrage and their eventual repeal. That pattern dates back at least to the creation of the Eurodollar market in the 1960s and the subsequent slow removal of deposit ceilings and many reserve requirements. If regulatory circumvention is not possible, in the longer term the firm can plan its organization and its capital and funding strategy to create more flexibility in managing regulatory requirements, usually at the cost of holding excess capital and liquidity in some units.

Therefore, for a given scale of its various businesses, there are regulatory costs that the firm can minimize to some extent. Once again, the firm will invest in information technology and management of regulatory requirements only if the diversification benefits (taking into account the ability to manage capital and liquidity on a very short-term basis under contingencies) are seen to exceed the information and regulatory costs. Moreover, the reduction of regulatory barriers to moving capital and liquidity within the firm enables the development or enlargement of the firm's internal capital market and increases the gains from pooling risk measurement information within the firm as well as the firm's overall efficiency.

Financial Condition

Intensive work on consolidated risk management has coincided with the rebuilding of the financial strength of many banking organizations following the difficulties of the late 1980s and early 1990s. For instance, a 1998 Federal Reserve study of credit risk models (Federal Reserve System Task Force on Internal Credit Risk Models 1998) notes that large U.S. banks have begun to develop both advanced credit risk modeling and internal capital allocation systems only since the mid-1990s—just the period over which these institutions recovered from the financial stresses of the earlier part of the decade. These internal capital allocation systems are one of the key elements in banks' attempts to evaluate the risk-adjusted performance of their various business units. As such, they represent an important step in the progress toward full-fledged consolidated risk management systems.[22]

This financial rebuilding may also have contributed to the growing emphasis on consolidated risk management systems. As argued above, one key motivation for consolidated risk management is to enable firms to make more informed judgments about where to invest their scarce capital resources, in particular, about where to expand through acquisition or internal growth. Firms in weakened financial condition are unlikely to be in a position to fund such growth—even into lines of business where the institution's risk/return trade-off is highly favorable—so they have less incentive to invest in the consolidated risk management systems that would permit them to identify such opportunities. The improved financial condition of many institutions since the early part of the 1990s therefore may have provided an additional incentive for firms to develop and implement consolidated risk management systems.

Intensive work on consolidated risk management has coincided with the rebuilding of the financial strength of many banking organizations following the difficulties of the late 1980s and early 1990s.

Declining information costs, eroding regulatory barriers, and stronger financial condition present fairly stylized explanations for increased attention by financial institutions to consolidated risk management and internal capital allocation activities. However, the optimization problem faced by firms is more complex than we have described. Holmstrom and Roberts (1998) provide many examples of the rich variety of mechanisms used to coordinate activities within and among firms and the multiplicity of factors that influence the coordination decision. The examples particularly illustrate the roles that incentives in internal (implicit) and external contracts and information flows play in resolving complex coordination problems, including overcoming regulatory barriers.[23] The implication is that coordination mechanisms used by individual firms may change as a wide variety of factors change. The current importance of consolidated risk management as a goal for many financial institutions could be enhanced or complemented by further advances in information technology and monitoring techniques, new designs for incentive contracts with employees and outside agents, better public and private information flows, and greater liquidity of financial markets.

Even so, the decline of information costs and the erosion and repeal of regulatory barriers have been so great that many of the principal hurdles to consolidated risk management within a financial conglomerate involve problems in measuring, comparing, and aggregating risks across business lines. The ability to merge banks and insurance companies under the Gramm-Leach-Bliley Act provides financial institutions with new opportunities to diversify risks and expand internal capital markets and creates further impetus to develop consolidated risk management techniques for financial conglomerates. Thus, both firms and supervisors are probably closer today in their common interest in accurate and precise risk measurement than they were just five years ago.

Major Technical Challenges and Research Questions

The previous sections discussed the economic rationale behind consolidated risk management and some of the costs facing diversified financial firms in constructing such systems. In this section, we turn to some additional practical problems associated with this overall goal. Our goal is to highlight a series of practical issues where additional research by risk management practitioners and by academics would be especially beneficial. In particular, we describe some of the technical challenges involved in actually estimating an aggregate measure of risk for a diversified financial institution and suggest some areas where further research could help both financial institutions and supervisors understand the strengths and weaknesses of such aggregate risk management.

At a very general level, there does appear to be an emerging consensus about how various forms of risk should be quantified. Most risk measurement methods used by major financial institutions are intended to capture potential losses over some future horizon due to the risk in question. These methods can use a probability-weighted approach to estimating potential losses (as in a value-at-risk or earnings-at-risk system, where the distribution of future earnings is calculated) or can provide point estimates of potential losses under certain extreme circumstances (as in a stress test or scenario analysis approach or in an "expected tail loss" estimation). The common thread is the focus on potential future losses, either to earnings or economic value.[24]

Beyond this general consensus, however, the picture is considerably more complex. As noted above, an aggregate risk measure must incorporate different types of risk (market, credit, operational) and must bring together risks across different business lines (banking, insurance, securities). Although the broad risk concept applied within and across these two dimensions may be similar, the details differ considerably, making simple "bottom-up" aggregation approaches difficult, if not impossible, to implement.

Aggregating across business lines presents challenges because firms and functional supervisors in the different business lines have tended to approach risk management and measurement from quite different perspectives. For instance, banks traditionally have emphasized the risks arising from the asset side of the balance sheet (credit risk) and from the interaction of assets, liabilities, and off-balance-sheet positions (interest rate risk, liquidity risk). Insurers, in contrast, have tended to place emphasis on the risks arising from the liability side of their business (underwriting risk, catastrophe risk). Securities firms have tended to emphasize the combination of market risk and liquidity (meaning both the ability to fund or to sell an asset) in their portfolios. Of course, advances in financial theory and market practice have eroded these distinctions somewhat, and many firms now attempt to measure the way in which risks can interact and affect an entire institution.[25] Nonetheless, one of the key challenges of consolidated risk management is to integrate these different perspectives on risk into a coherent framework.

> *Our goal is to highlight a series of practical issues where additional research by risk management practitioners and by academics would be especially beneficial.*

A related set of challenges arises when aggregating across different types of risk. These challenges reflect the fact that at many financial institutions, risk measurement and management began as a bottom-up process, with different types of risk measured separately. A particular business area would develop risk measurement approaches to capture the most important risks facing that unit: credit risk for lending activities, market risk for trading, interest rate risk for the treasury/asset-liability management function. This risk-by-risk approach has resulted in industry standards of risk measurement that differ significantly across risk types, and sometimes across activities with similar risks, both in the way that risk is measured and in the extent to which it is quantified at all.

To a large extent, the state of development of modeling technology across the various risks reflects the availability of data and the nature of the risk itself, which can affect the ease or difficulty involved in accurately modeling the risk. At one end of the spectrum, the banking and securities industry has a now fairly long history of measuring market risk through value-at-risk models. The fact that value-at-risk models were among the first statistical risk models developed reflects the high-frequency and largely continuous nature of market risk and its management,[26] the mark-to-market environment in which most trading activities occur, and the resultant ease of modeling (normality has often been assumed) and availability of comparatively long historical data series around which to calibrate the models.

Credit risk tends to exhibit somewhat lower frequency variation, as changes in credit status tend to evolve over weeks or months rather than on a day-to-day basis. Thus, fewer historical data are available to aid in model calibration, and the modeling process itself is more complex, as the distribution of credit losses is quite asymmetric with a long right-hand tail.[27] Financial institutions have made considerable progress over the past two or three years in credit risk modeling, but it is fair to say that these models are at an earlier stage of development than the value-at-risk models used for market risk assessment.[28]

At many financial institutions, risk measurement and management began as a bottom-up process, with different types of risk measured separately.

Even further down the spectrum is operational risk—the risk stemming from the failure of computer systems, control procedures, and human error—which captures a mixture of events, some of which involve relatively frequent small losses (settlement errors in a trading operation, for instance) and others that are characterized by infrequent but often large losses (widespread computer failure). Consistent data sources on this form of risk are difficult to obtain, especially for the less frequent events; statistical modeling is in its early stages; and the computational requirements may be substantial, given the number of "control points" in most operational processes.

Liquidity risk measurement involves many similar issues of sorting the frequency of different types of events and developing appropriate data. Liquidity risk measurement has long involved scenario analysis focused on stress events and based on subjective probabilities of how depositors, other creditors, and borrowers would respond to the stress event. As risk measurement techniques have advanced, some financial institutions are examining the potential for cash-flow-at-risk analysis, based on more formal measurement of the probability of events and the sensitivity of cash flows to these events, both to enhance day-to-day liquidity management and to strengthen the underpinnings of liquidity stress scenarios. Finally, at the far end of the spectrum, other risks—such as legal, reputational, and strategic risk—are rarely quantified, as both the data and theoretical techniques for capturing these risks have yet to be developed extensively.

Even for those risks that are measured, important differences exist in the assumptions and techniques used to estimate potential losses. One key issue is the time horizon over which potential losses are to be measured. As noted above, the risks facing financial institutions vary in the extent to which they are continuous or discrete, in how quickly new events develop, and in the size of events when they occur (many small events versus a few large ones). These differences imply the need for different horizons to capture different risks effectively. In fact, we see these differences in the assumptions underlying the risk estimates made by financial firms, with market risk typically measured over a one-day horizon, credit risk typically measured over a one-year horizon, and operational risk measured over a variety of short and long horizons (an industry standard has yet to emerge).

These differences present a challenge for calculating consolidated risk exposures that span several risk types. Should a single horizon be chosen for all risks and, if so, which one? Should the time dimension be explicitly factored into the risk assessment, with paths of risk over time? More generally, issues such as differing horizons suggest that there is an important set of research questions concerning methods for calculating aggregate risk measures. At a very basic level, can the different individual risk measurement approaches typically used within financial firms be meaningfully aggregated? If so, how? If not, is it possible to develop a "top-down" approach that somehow blends the risks facing the firm without measuring them separately, such as an analysis of income volatility? Is there some way of combining "top-down" with "bottom-up" approaches to consolidated risk measurement? And how does the growing attention to evaluating performance against risk in rewarding managers at all levels of the organization factor into these decisions?

Perhaps a more fundamental question is whether a consolidated risk management *system needs to have a fully consolidated risk* measurement *methodology at its core.*

A related set of issues concerns the mathematical aggregation of risk measures across businesses and risk types. In most cases, this process would involve estimating correlations between various risk exposures. An important challenge in this regard is measuring the degree of correlation between risks in businesses that are distinct in terms of the sources and frequency of variability (for instance, between insurance underwriting and trading). The data demands of producing accurate estimates are likely to be enormous. Even when aggregate risk measures can be calculated, a related challenge is how to apportion the benefits of diversification across various business lines. That is, if less-than-perfect correlation across distinct business lines results in a decrease in the overall risk facing the firm, how should these benefits be allocated back to the various business units in the internal capital allocation process?

This discussion assumes that to produce a consolidated measurement of risk exposure, it is necessary to develop risk measures that are highly comparable across risk types. However, perhaps a more fundamental question is whether a consolidated risk *management* system needs to have a fully consolidated risk *measurement* methodology at its core. In other words, how much comparability across risk measures is strictly necessary to have an effective consolidated risk management system? If risk measures cannot be made perfectly compatible across risk types and business lines, are there still benefits to imperfectly comparable measures?

Our sense is that the answer to this question is likely to be a resounding yes, largely because the ability to evaluate results against risks taken has become a major feature of financial institution management in the 1990s. Some important issues would need to be explored before understanding the full implications of this conclusion. For instance, what kind of biases might enter the assessment of aggregate risk if this assessment is based on disparate risk measures? How might comparisons of risk and return across business lines be affected? How can we relate the results of stress scenario analysis to statistical measures of risk exposure? Are there limits to how different the various risk measures can be, yet still be useful in a consolidated risk management system? These are important, unresolved issues.

Conclusion

As the above discussion suggests, there is considerable scope for further research to enhance our understanding of the benefits and shortcomings of consolidated risk management. Many of the key research questions involve technical issues in risk measurement and financial series modeling. While these questions are vital to understanding how to calculate a consolidated measure of risk exposure spanning all of a financial institution’s businesses and risk factors, they are not the only questions of interest. Further research into the main question of this article—the economic rationale for consolidated risk management—could produce findings that would be of clear use to supervisors and financial institutions. In addition, this work could provide insight into such diverse topics as the theory of the firm, the costs/benefits of diversification, the linkages among financial markets, and the impact of product and geographic deregulation. Our study presents some initial ideas, but clearly much more work needs to be done. We hope that this article can serve as a starting point for further discussion.

Endnotes

1. It can also be argued that supervisors may place somewhat greater weight on the risk of severe downside scenarios, given the nature of the supervisory role, but the private sector appears to be closing any gap as a result of the insight gained from experiences such as the market disturbances in 1998.

2. Firms vary in how they use the risk management process to maximize profits. Some firms use risk-and-return measures in the selection of their medium-term business mix in order to maximize long-run expected profits. Firms also use risk management systems to assist in managing expected profits over short horizons, by seeking to identify changes in risk and loss potential and adjusting their portfolios accordingly.

3. In large measure, these efforts are an extension of a longer term trend toward enhanced risk management and measurement in the financial services industry. Many of these efforts have focused on developing risk measurement and management systems for individual risk types or businesses (for instance, market risk in a securities firm or credit risk in a bank's loan portfolio). In consolidated risk management, however, the focus is on an expansion of these single-risk-management systems to span diverse financial activities, customers, and markets.

4. Mergers may occur for many reasons, including the desire to benefit from exactly the sort of diversification that presents challenges to risk management and measurement systems. In this discussion, we distinguish between the broad diversification that may occur when firms comprise business units involved in distinct business activities (such as banking, insurance, or securities activities) or geographic locations and the type of portfolio diversification that occurs when risk management units take steps to hedge portfolio- or business-level risk exposures. It is the first type of diversification—which has become much more feasible given the regulatory and technical developments discussed in the text—that presents the sort of challenges we discuss in this article.

5. The evaluation of the adequacy of risks in light of a full risk assessment is discussed in Federal Reserve SR Letter 99-18. Earlier in the decade, the Federal Reserve issued SR Letter 93-69, on the management of trading activities; SR Letter 97-32, on information security; SR Letter 00-04, on outsourcing; and a series of papers on the management of credit risk in both primary and secondary market activities (SR Letters 99-3, 98-25, and 97-21). The Office of the Comptroller of the Currency and the Federal Deposit Insurance Corporation have issued guidance using a comparable framework for a similar range of topics.

6. This framework is best developed in "Principles for the Management of Credit Risk," published in September 2000. The Committee has also published work on interest rate risk, in 1997; operational risk, in 1998; and liquidity risk, in 2000.

7. The work of Modigliani and Miller (1958) and Miller and Modigliani (1961) suggests that any risk-altering actions taken by a firm's management are redundant and resource-wasting because shareholders can achieve their optimal degree of diversification independently. See Cummins, Phillips, and Smith (1998) for a discussion of the factors—such as bankruptcy costs, taxes, and costly external financing—that may make it worthwhile for firms to engage in risk management.

8. This relationship can be expressed mathematically as

$$\sigma_{FIRM} = \sqrt{\sigma_A^2 + \sigma_B^2 + 2\rho\sigma_A\sigma_B} \leq \sigma_A + \sigma_B ,$$

where σ_A and σ_B are the profit volatilities of business units A and B and ρ is the correlation between them.

9. This situation was not uncommon among globally dispersed institutions prior to the introduction of enhanced information systems in the early-to-mid-1990s. Later in this article, we discuss the role of information costs and information systems in diversified financial institutions.

10. Morris and Shin (1999) describe this problem in the context of multiple firms operating in a single market. They describe the errors in risk assessment that can occur when risk management systems assume that the firm's activities are similar to playing roulette (gambling against nature), when in fact the risks are more like those in poker (where the actions of the other players matter). The same analogy can be applied to risks within a firm.

11. Or, as discussed below, such contagion fears may arise because market observers believe that the resources of all of the firm's business units will be used to "rescue" a troubled unit, calling into question the solvency of all of the businesses within the firm.

12. The large investments that many financial institutions are making in electronic trading and banking are examples of strategic risk related to establishing the competitive position of a firm in a fast-changing and greatly contested market. The problems many financial institutions experienced in the mid-1990s—when customers experienced large losses in connection with derivatives and complex trading strategies—are examples of strategic risk related to damage to the firm's reputation.

13. Froot and Stein (1998) consider a variant of this risk—the bankwide cost of capital effect—that involves the impact of increased capital costs on all units within a firm when one unit takes on large amounts of risk.

14. In the Froot and Stein analysis, banks choose their capital structure, risk management policies, and investment policies jointly, rather than impose a short-run capital constraint. However, when capital is costly, banks economize on the amount of capital they hold and therefore take risk management concerns into account in their investment policies.

15. The example Froot and Stein give is the counterparty risk on a foreign exchange swap. With the advent of credit derivatives and other credit risk management techniques, such risks are increasingly tradable, by which Froot and Stein mean that the risks can be offset to achieve a zero net present value. Nontradable risks can include unhedged proprietary positions premised on subjective expected rates of return deviating from those of the market. Note that the reliance on markets for hedging for liquid risks and internal capital allocation for nontradable risks is another version of the contractible/noncontractible distinction discussed earlier.

16. Other creditors here could include suppliers, consultants, and other contractors who provide products or services in return for the promise of future payment.

17. This would be especially true if there were meaningful disclosures about intrafirm exposures, as called for in a recent report by the Joint Forum (1999b).

18. As information systems become more "scalable," the step function may become flatter, in effect making it easier to realize and manage the diversification benefits from combining activities.

19. This is also consistent with the analysis of Holmstrom and Milgrom (1994), which derives analytically that enhancements to performance measurement tend to permit greater employee freedom (such as higher limits), although the authors caution that their analysis requires a careful specification of the exact problem.

20. Correlations and volatilities have changed substantially over time. Examples include the sharp drop in volatilities in short-term interest rates associated with the decline in inflation in the 1980s and early 1990s; sharp increases in the correlations and short-term volatilities of U.S. long-term fixed income instruments in times of distress; and a rise in the idiosyncratic risk of equities in the 1990s, the last example documented in Campbell, Lettau, Malkiel, and Xu (2001).

21. Gibson (1998) derives similar conclusions about the impact of declining information costs. In his approach, risk measurement is a means to monitor risk-taking by employees when information about the managerial effort of those employees (or outside agents, such as mutual fund managers) is not observable by the employer.

22. Typically, these internal capital allocation systems fall short of a full-fledged consolidated risk measurement system, either because they incorporate only a limited range of the risks facing a financial institution (for example, just credit risk or market risk, but not operational or other forms of risk) or because they are applied only to a subset of the institution's activities.

23. The specific regulatory barrier they cite is the limitations on foreign ownership of domestic airlines.

24. Other potential definitions of risk could involve pure volatility measures, such as standard deviations of earnings or economic value, or sensitivity measures that capture the derivative of earnings or economic value with respect to particular risk factors, such as the "value of a basis point."

25. Lewis (1998), for instance, describes how one insurance company examines stress scenarios that affect all aspects of the firm, such as an earthquake that simultaneously causes extremely high insurance claims and disrupts financial markets—and thus the firm's investments and investment income—for some period of time.

26. Of course, some market price series exhibit sharp, discontinuous jumps, such as those associated with emerging market developments and unexpected changes in exchange rate regimes. These factors have tended to be incorporated into value-at-risk models after the initial phases of model development.

27. To some extent, both the lack of data and the lower frequency variation reflect the current GAAP (Generally Accepted Accounting Principles) accounting standards, which do not require the daily marking-to-market to which trading account positions are subject. Thus, shorter term variation in value may not be reflected in the accounting data typically available for use in credit risk models.

28. See Basel Committee on Banking Supervision (1999a) for a discussion of the state of development of credit risk models.

References

Basel Committee on Banking Supervision. 1999a. "Credit Risk Modeling: Current Practices and Applications." April. Basel, Switzerland: Bank for International Settlements.

———. 1999b. "A New Capital Adequacy Framework." June. Basel, Switzerland: Bank for International Settlements.

———. 1999c. "Performance of Models-Based Capital Charges for Market Risk." September. Basel, Switzerland: Bank for International Settlements.

———. 2000a. "Sound Practices for Managing Liquidity in Banking Organizations." February. Basel, Switzerland: Bank for International Settlements.

———. 2000b. "Principles for the Management of Credit Risk." September. Basel, Switzerland: Bank for International Settlements.

Board of Governors of the Federal Reserve System. 1993. "Examining Risk Management and Internal Controls for Trading Activities of Banking Organizations." SR Letter 93-69. December 20.

———. 1997a. "Risk Management and Capital Adequacy of Exposures Arising from Secondary Market Credit Activities." SR Letter 97-21. July 11.

———. 1997b. "Sound Practice Guidance for Information Security for Networks." SR Letter 97-32. December 4.

———. 1998. "Sound Credit Risk Management and the Use of Internal Credit Risk Ratings at Large Banking Organizations." SR Letter 98-25. September 21.

———. 1999a. "Supervisory Guidance Regarding Counterparty Credit Risk Management." SR Letter 99-3. February 1.

———. 1999b. "Assessing Capital Adequacy in Relation to Risk at Large Banking Organizations and Others with Complex Risk Profiles." SR Letter 99-18. July 1.

———. 2000. "Outsourcing Information and Transaction Processing." SR Letter 00-04. February 29.

Campbell, John, Martin Lettau, Burton Malkiel, and Yexuai Xu. 2001. "Have Individual Stocks Become More Volatile: An Empirical Exploration of Idiosyncratic Risk." Journal of Finance 56, no. 1 (February): 1-43.

Coase, Ronald. 1937. "The Nature of the Firm." Economica 4, no. 4 (November): 386-405.

Counterparty Risk Management Policy Group. 1999. "Improving Counterparty Risk Management Practices." June.

Cummins, J. David, Richard D. Phillips, and Stephen D. Smith. 1998. "The Rise of Risk Management." Federal Reserve Bank of Atlanta Economic Review, first quarter: 30-40.

Federal Reserve System Task Force on Internal Credit Risk Models. 1998. "Credit Risk Models at Major U.S. Banking Institutions: Current State of the Art and Implications for Assessments of Capital Adequacy." May.

Froot, Kenneth A., and Jeremy C. Stein. 1998. "Risk Management, Capital Budgeting, and Capital Structure Policy for Financial Institutions: An Integrated Approach." Journal of Financial Economics 47, no. 1 (January): 55-82.

Gertner, Robert H., David S. Scharfstein, and Jeremy C. Stein. 1994. "Internal versus External Capital Markets." Quarterly Journal of Economics 109, no. 4 (November): 1211-30.

Gibson, Michael S. 1998. "The Implications of Risk Management Information Systems for the Organization of Financial Firms." Board of Governors of the Federal Reserve System International Finance Discussion Paper no. 632, December.

Global Derivatives Study Group. 1993. "Derivatives: Practices and Principles." July. Washington, D.C.: Group of Thirty.

Grossman, Sanford J., and Oliver D. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." Journal of Political Economy 94, no. 4: 691-719.

Holmstrom, Bengt, and Paul Milgrom. 1994. "The Firm as an Incentive System." American Economic Review 84, no. 4 (September): 972-91.

Holmstrom, Bengt, and John Roberts. 1998. "The Boundaries of the Firm Revisited." Journal of Economic Perspectives 12, no. 4 (fall): 73-94.

References (Continued)

Joint Forum (Basel Committee on Banking Supervision, International Organization of Securities Commissioners, International Association of Insurance Supervisors). 1999a. "Risk Concentration Principles." December. Basel, Switzerland: Bank for International Settlements.

———. 1999b. "Intra-Group Transactions and Exposures Principles." December. Basel, Switzerland: Bank for International Settlements.

Labrecque, Thomas G. 1998. "Risk Management: One Institution's Experience." Federal Reserve Bank of New York Economic Policy Review 4, no. 3 (October): 237-40.

Lam, James. 1999. "Enterprise-Wide Risk Management: Staying Ahead of the Convergence Curve." Journal of Lending and Credit Risk Management 81, no. 10 (June): 16-9.

Lewis, Robert. 1998. "Capital from an Insurance Company Perspective." Federal Reserve Bank of New York Economic Policy Review 4, no. 3 (October): 183-5.

Miller, Merton H., and Franco Modigliani. 1961. "Dividend Policy, Growth, and the Valuation of Shares." Journal of Business 34 (October): 411-33.

Modigliani, Franco, and Merton H. Miller. 1958. "The Cost of Capital, Corporation Finance, and the Theory of Investment." American Economic Review 48 (June): 261-97.

Morris, Stephen, and Hyun Song Shin. 1999. "Risk Management with Interdependent Choice." Bank of England Financial Stability Review, no. 7 (November): 141-50.

Mudge, Dan. 2000. "The Urge to Merge." Risk Magazine. February, p. 64.

Beverly J. Hirtle, Mark Levonian, Marc Saidenberg, Stefan Walter, and David Wright

Using Credit Risk Models for Regulatory Capital: Issues and Options

- Regulatory capital standards based on internal credit risk models would allow banks and supervisors to take advantage of the benefits of advanced risk-modeling techniques in setting capital standards for credit risk.
- The internal-model (IM) capital standards for market risk provide a useful prototype for IM capital standards in the credit risk setting.
- Nevertheless, in devising IM capital standards specific to credit risk, banks and supervisors face significant challenges. These challenges involve the further technical development of credit risk models, the collection of better data for model calibration, and the refinement of validation techniques for assessing model accuracy.
- Continued discussion among supervisors, financial institutions, research economists, and others will be key in addressing the conceptual and theoretical issues posed by the creation of a workable regulatory capital system based on banks' internal credit risk models.

In January 1996, the Basel Committee on Banking Supervision adopted a new set of capital requirements to cover the market risk exposures arising from banks' trading activities. These capital requirements were notable because, for the first time, regulatory minimum capital requirements could be based on the output of banks' internal risk measurement models. The market risk capital requirements thus stood in sharp contrast to previous regulatory capital regimes, which were based on broad, uniform regulatory measures of risk exposure. Both supervisors and the banking industry supported the internal-models-based (IM) market risk capital requirement because firm-specific risk estimates seemed likely to lead to capital charges that would more accurately reflect banks' true risk exposures.

That market risk was the first—and so far, only—application of an IM regulatory capital regime is not surprising, given the relatively advanced state of market risk modeling at the time that the regulations were developed. As of the mid-1990s, banks and other financial institutions had devoted considerable resources to developing "value-at-risk" models to measure the potential losses in their trading portfolios. Modeling efforts for other forms of risk were considerably less advanced. Since that time, however, financial institutions have made strides in developing statistical models for other sources of risk, most notably credit risk. Individual banks have developed proprietary models to capture potential credit-related losses from their loan portfolios, and a variety of models are available from consultants and other vendors.

Beverly J. Hirtle is a vice president at the Federal Reserve Bank of New York, Mark Levonian is a director in the Banking Supervision and Regulation Division at the Federal Reserve Bank of San Francisco, Marc Saidenberg is a Bank Supervision officer and Stefan Walter a vice president at the Federal Reserve Bank of New York, and David Wright is an assistant director of the Banking Supervision and Regulation Division at the Board of Governors of the Federal Reserve System.

The authors would like to thank Edward Ettin, Michael Gordy, Darryll Hendricks, David Jones, Jose Lopez, Brian Peters, and two anonymous referees for many thoughtful comments. The views expressed are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York, the Federal Reserve Bank of San Francisco, or the Federal Reserve System.

These developments raise the question of whether banks' internal credit risk models could also be used as the basis of regulatory minimum capital requirements. The Basel Committee on Banking Supervision is in the midst of revising regulatory capital standards and has in fact considered using credit risk models for this purpose. However, in a study released in April 1999 (Basel Committee on Banking Supervision 1999a), the Committee concluded that it was premature to consider the use of credit risk models for regulatory capital, primarily because of difficulties in calibrating and validating these models.

The purpose of this article is to build on this earlier work, by the Basel Committee and others, and to consider the issues that would have to be addressed in developing a regulatory minimum capital standard based on banks' internal credit risk models. In conducting this exercise, we consider how such a capital regime might be structured if the models were sufficiently advanced. This article is *not* intended to be a policy proposal, but rather to serve as a discussion laying out the issues that would have to be addressed in creating a capital framework based on credit risk models. In particular, we draw on the structure of the IM capital charge for market risk and examine how this structure might be applied in the credit risk setting.

As in the market risk setting, the overall objective of an internal-models regulatory capital charge would be to allow banks and supervisors to take advantage of the benefits of advanced risk-modeling techniques in setting capital standards for credit risk. Ideally, the framework should provide supervisors with confidence that the IM capital charges are conceptually sound, empirically valid, and reasonably comparable across institutions. At the same time, an IM framework should be flexible enough to accommodate—and perhaps even encourage—further innovation in credit risk measurement. The balance between meeting immediate prudential needs and fostering continuing, fruitful innovation is one of the key themes in the discussion that follows.

The overall objective of an internal-models regulatory capital charge would be to allow banks and supervisors to take advantage of the benefits of advanced risk-modeling techniques in setting capital standards for credit risk.

The remainder of this article lays out the issues that would be involved in structuring an IM capital regime for credit risk exposures. The next section contains a brief overview of the basic concepts underlying credit risk models. We then describe the basic components of an IM capital framework for credit risk—prudential standards, modeling standards, and validation techniques—and discuss a range of alternative approaches for these standards. At certain points in this discussion, we identify particularly difficult issues that would have to be addressed before an IM framework could be implemented. In such cases, we describe the scope of the issues and their importance, rather than make specific recommendations.

Overview of Credit Risk Models

This section provides a brief overview of credit risk models.[1] The purpose of this discussion is to provide background about the general structure and key features of credit risk models that will help explain the regulatory capital framework described in the next section. For this purpose, we will focus on the concepts that are common to all credit risk models, rather than present a detailed description of specific models. It is also important to note that the models described in this section are those that are usually applied to banks' wholesale and middle-market commercial lending portfolios. The models used for some other types of credits—for example, retail lending such as credit cards, auto loans, and small business loans—generally differ from the models described below.

In very general terms, the purpose of a credit risk model is to estimate the probability distribution of future credit losses on a bank's portfolio. The first step in constructing a credit risk model is therefore to define the concept of loss that the model is intended to capture, as well as the horizon over which the loss is measured. In terms of the definition of loss, models generally fall into one of two categories: models that measure the losses arising solely from defaults ("default mode" models), and models that incorporate gains and losses arising from less extreme changes in credit quality as well as from defaults ("multistate" or "mark-to-market" models). Clearly, the default mode paradigm is a restricted version of the multistate approach, and some models are designed to produce loss estimates based on both definitions of loss.

For both approaches, losses are measured over some future planning horizon. The most common planning horizon used is one year, meaning that the model will estimate changes in portfolio value—either from defaults or from more general changes in credit quality—between the current date and one year in the future. While a one-year horizon is most common in practice, other choices are also possible, including fixed horizons other than one year and horizons that match the lifetime of the credits in the portfolio.

Once the definition of loss and the planning horizon have been selected, the model generates a distribution—a probability density function (PDF)—of future losses that can be used to calculate the losses associated with any given percentile of the distribution. In practice, banks concentrate on two such loss figures: *expected loss* and *unexpected loss*. Expected loss is the mean of the loss distribution and represents the amount that a bank expects to lose on average on its credit portfolio. Unexpected loss, in contrast, is a measure of the variability in credit losses, or the *credit risk* inherent in the portfolio. Unexpected loss is computed as the losses associated with some high percentile of the loss distribution (for example, the 99.9th percentile) minus expected loss. A high percentile of the distribution is chosen so that the resulting risk estimates will cover all but the most extreme events.

The first step in generating the PDF of future credit losses is to classify the individual credits in the portfolio by their current credit quality. Most frequently, this is done by distributing the credits across the bank's internal credit risk rating system, which provides a picture of the current state of the credit portfolio. Typically, a bank will have an internal rating system that assigns each credit to one of a series of risk categories according to the borrower's probability of default. The next conceptual step is to assess the probability that the positions might migrate to different risk categories—sometimes called "credit quality states"—during the planning horizon. In a default mode model, this process amounts to assessing the probability of default, while in a multistate model, it also incorporates assessing transition probabilities between internal rating categories. The accuracy of both the assignment and the quantification of banks' internal risk ratings is critical, as these ratings and transition probabilities have a very significant effect on the estimation of portfolio credit risk.[2]

The third step in constructing a credit risk model is to estimate the likely exposure of each credit across the range of credit quality states. For whole loans, exposure is simply the face value of the loan and is usually constant across risk categories, but for other positions—such as lines of credit or derivatives—exposure can vary over time and might be correlated with the particular credit quality state. Finally, given the risk category and the exposure in that category, the last element to be determined is the valuation of the position. For default mode models, this valuation is usually accomplished by specifying a loss-given-default (LGD) percentage. This is, essentially, the proportion of the credit's exposure that would be lost if the borrower defaults.[3] For multistate models, this process generally involves revaluing the position using credit spreads that reflect the default risk associated with the particular rating category.

Thus far, the discussion has focused on the treatment of individual positions in a bank's credit portfolio. Generating the PDF of future credit losses requires bringing these individual positions together to capture the behavior of the overall portfolio. From standard portfolio theory, this process essentially requires capturing the correlations between losses associated with individual borrowers. Correlations are vital in assessing risk at the portfolio level since they capture the interaction of losses on individual credits. In general, portfolio risk will be greater the more the individual credits in the portfolio tend to vary in common. In practice, incorporating correlations into a credit risk model involves capturing variances in and correlations between the risk category transition probabilities, credit exposures, and credit valuations.

Nearly all models assume that these variances and correlations are driven by one or more "risk factors" that represent various influences on the credit quality of the borrower (for example, industry, geographic region, or the general state of the economy). In some models, risk factors are economic variables such as interest rates and economic activity indicators, while other models derive default and transition probabilities from equity price data. In still other models, the risk factors are abstract factors that intuitively relate to business cycle conditions but are not tied to specific economic variables. In every case, the assumptions about the statistical process driving these risk factors determine the overall mathematical structure of the model and the shape of the PDF.[4] Thus, assumptions about the distribution of risk factors are a key element in the design of all credit risk models.

> *Assumptions about the distribution of risk factors are a key element in the design of all credit risk models.*

Depending on the assumptions about the mathematical processes driving the risk factors, there are a variety of ways that the final PDF of future credit losses can be generated. In some cases, a specific functional form for the PDF is assumed and the empirical results are calculated analytically. In other cases, Monte Carlo simulation—generally involving simulation of the underlying risk factors that determine default and transition probabilities—is used to provide a numerical PDF. In either case, the final result is a PDF that can be used to derive estimates of the various percentiles of the loss distribution.

Framework for an Internal-Models Capital Charge

This section describes a possible framework for an internal-models regulatory capital charge for credit risk exposures. In developing this framework, we use the IM capital requirements for market risk as a model.[5] As a practical matter, the market risk standards provide a foundation that would be familiar to the many parties involved in developing and implementing any new credit risk standards. On a theoretical level, it also seems reasonable to use the market risk framework as a starting point because, fundamentally, both market and credit risk models have the same goal: to estimate the distribution of gains and losses on a bank's portfolio over some future horizon. The two types of models differ with respect to the underlying risk factors that generate these gains and losses, and these differences lead to significant differences in methodologies, modeling assumptions, and data requirements between the models. Nonetheless, the core similarity between the two types of models suggests that the framework used in the market risk setting can provide a workable beginning for a regulatory capital regime based on internal credit risk models.

> *Perhaps the most appealing approach would be to base an internal-models regime on a* multistate loss concept*, because it takes account of the probability of changes in credit quality as well as the probability of default.*

As noted above, the basis of the market risk requirements is a risk measurement model that estimates the distribution of gains and losses on the bank's portfolio over some future time horizon. The market risk capital charge is based on a certain percentile of this distribution. In particular, the capital charge is based on the 99th percentile loss amount over a ten-day future time horizon. This amount represents the maximum that the bank could lose over a ten-day period with 99 percent probability. Such estimates are often interpreted as measures of the degree of risk inherent in a bank's portfolio, since they reflect the portfolio's potential for future losses.

A regulatory capital requirement for credit risk could be based on the output of credit risk models in a similar fashion. Just as in the market risk setting, the capital charge could be based on a particular percentile of this loss distribution over a given time horizon. These parameters would differ from those used in the market risk capital framework, for reasons that are discussed below. Nonetheless, the basic structure of the framework—a capital requirement based on a statistical estimate of the distribution of future gains and losses on the bank's positions—could be applied to credit risk exposures.

As in the market risk setting, the IM framework for credit risk could have three general components: a set of prudential standards defining the risk estimate to be used in the capital charge, a set of model standards describing the elements that a comprehensive credit risk model would incorporate, and validation techniques that could be used by supervisors and banks to ensure that model estimates are reasonably accurate and comparable across institutions. These three general components could be specified in a variety of ways, and the discussion that follows generally highlights a range of alternatives. The goal of the discussion is to provide a sense of the features that an IM approach to regulatory capital would likely incorporate and to raise issues requiring further analysis and comments.

Prudential Standards

The first component of an IM regulatory capital regime would be a set of prudential standards intended to establish the basic degree of stringency of the capital charge. As such, these standards would be specified by the supervisor to ensure that the regulatory capital requirements provide a suitable degree of prudential coverage and would be the same for all banks subject to the capital charge. Mirroring the basic elements of credit risk measurement models described in the previous section, these prudential standards would include the definition of loss, the planning horizon, and the target loss percentile. Each of these elements is discussed below.

Definition of Loss

As noted, the first step in specifying a credit risk model is to determine the definition of loss and the planning horizon. Similarly, in constructing a minimum capital requirement based on internal models, the first step would be to specify supervisory standards for these concepts. In particular, an IM approach to regulatory capital would need to specify whether the minimum capital requirement would be based on a default mode or multistate loss concept and the horizon over which these losses would be measured.

From a prudential perspective, the two standards are linked, since there is something of a trade-off between the length of the planning horizon and the definition of loss. Specifically, longer planning horizons appear appropriate for the default mode approach since the impact of defaults that occur beyond the end of the planning horizon is ignored. Conversely, somewhat shorter planning horizons may be acceptable in a multistate paradigm because some of the impact of these long-term defaults is captured by credit rating downgrades.

Perhaps the most appealing approach would be to base an internal-models regime on a *multistate loss concept*, because it takes account of the probability of changes in credit quality as well as the probability of default. This approach is appealing because it recognizes economic gains and losses on the credit portfolio and, from a supervisory perspective, it holds the promise of requiring additional capital for credit weaknesses well in advance of their full development as losses. In addition, this approach is consistent with the growing tendency of many of the largest banking institutions to treat credit risk as something that can be traded and hedged in increasingly liquid markets. These considerations suggest that a multistate loss definition would be the soundest basis for a regulatory capital regime based on internal credit risk models.

Nonetheless, this choice would raise some issues that are worth noting. The most significant of these is that many models currently used by banks incorporate a default mode approach, which means that these models would have to be changed—and in some cases, entirely reconstructed—to be eligible for regulatory capital treatment. In addition, default mode models correspond in straightforward ways with the book value accounting used by many financial institutions, while multistate models are more consistent with market-value accounting. Thus, although some evidence suggests that the trend in the industry is moving away from default mode models and toward multistate approaches, the question remains whether a regulatory standard based on a multistate approach would place a significant burden on banks or whether it would merely provide them with the incentive to move more quickly in the direction that they were already going.

> *A* fixed one-year horizon *may represent the most workable balance between prudential concerns and practical considerations about modeling practice.*

Planning Horizon

As indicated above, the choice of a supervisory planning horizon is very much linked to the definition of loss. We have argued that a multistate loss definition that recognizes changes in credit quality short of default would provide the soundest basis for an IM capital regime for credit risk. Given this choice, we now consider several alternative planning horizons, including a fixed horizon of one year, a fixed horizon of more than one year, and a "lifetime" horizon that would cover the maturity of credits in a bank's portfolio.

At one end of the spectrum, a lifetime horizon would be consistent with the conceptual approach to a traditional banking book in which credits are held to maturity.[6] By looking over the full maturity of positions in the portfolio, the potential for all future losses would be captured by the capital requirement. In that sense, the lifetime assumption can be interpreted as requiring that capital be sufficient to ensure that, with a certain probability, the bank will be able to absorb any and all losses, even if it is unable to raise additional capital or to mitigate its troubled credits.

For this reason, the lifetime horizon would provide a very high degree of comfort that capital would be able to withstand quite significant negative credit events. However, the lifetime horizon approach is at odds with the modeling techniques in current use by most practitioners. In addition, the "buy and hold" portfolio management assumption might be excessively conservative in an environment in which credit risk is increasingly liquid. It seems likely, for instance, that even in stressful market situations, banks would have some ability to manage their loss exposures or to raise additional capital.

An intermediate approach to the loss horizon question might be to use a fixed horizon of several years. Since it can take two to three years (or longer) to work through the effects of a credit cycle, a fixed horizon of more than a year might be appropriate from a prudential perspective. However, few models currently incorporate a horizon of more than one year, so the benefits of increased prudential coverage would have to be weighed against the costs of altering the modeling approach most commonly used by banks.

For a variety of reasons, a *fixed one-year horizon* may represent the most workable balance between prudential concerns and practical considerations about modeling practice. As noted above, the multistate setting reflects the possibility of defaults beyond one year through credit downgrades during the year. Further, a one-year horizon may be sufficient for banks and supervisors to begin to respond to emerging credit problems. Finally, this horizon is consistent with market practice, and is the most commonly used approach in the industry. Thus, adopting a one-year horizon for regulatory capital purposes would be least disruptive to current modeling practice. This consideration—along with the fact that reasonable theoretical arguments can be constructed for different holding period assumptions—suggests that a one-year standard may be the most pragmatic approach.[7]

Target Loss Percentile

Along with the definition of loss and the planning horizon, the target loss percentile is one of the key prudential parameters of an internal-models-based regulatory capital regime. As in the market risk setting, the capital charge could be calculated based on the level of losses at *a specified percentile of the loss distribution*, minus the expected loss.[8] The specified percentile should be chosen so that, in conjunction with other parameters, the capital charge would provide the level of prudential coverage desired by the supervisory authorities.[9]

A number of considerations would apply in determining the appropriate target loss percentile. First, since the purpose of regulatory capital requirements is to ensure that banks hold sufficient capital to withstand significant losses, it seems reasonable to expect that the target loss percentile would be fairly high. For instance, those banks that use credit risk models for internal capital allocation purposes tend to pick target insolvency rates consistent with senior debt ratings in the mid-to-high investment-grade range. Historical data suggest that annual insolvency rates associated with such bonds are less than 1 percent, implying a target percentile above the 99th.[10] This example suggests that one approach to determining a target percentile is to consider the desired public debt rating for large banking institutions.

While safety concerns may suggest setting a very high target percentile, other considerations offset this incentive to some degree. First, the capital guidelines are meant to be minimum regulatory standards, and banks would almost certainly be expected to hold actual capital amounts higher than these minimums.[11] If this is the case, then it would be desirable to establish regulatory minimum capital requirements that are lower than the internal capital amounts that safe and prudent banks choose to hold.[12] This consideration suggests selecting a somewhat lower percentile of the distribution, perhaps one associated with the minimum public debt rating consistent with a bank's operating in a safe and sound manner.

There may also be practical reasons to consider selecting a somewhat lower target percentile. Foremost among these are validation issues. Since we observe losses associated with these high percentiles very infrequently, selecting a very high percentile as the supervisory standard may exacerbate the already difficult task of model validation. One possibility might be to base the regulatory capital requirement on a less extreme value of the PDF—for instance, the 90th percentile—that could be validated more easily and to adjust this figure upward if there is concern about whether the resulting capital charge was stringent enough. While this approach has certain intuitive appeal, establishing a scaling factor that would accurately translate a lower percentile loss estimate into the higher percentile desired for prudential reasons would require making parametric assumptions about the shape of the tail loss distribution. Given the lack of consensus among practitioners and researchers on this issue, as well as possible variation in the loss distribution across different types of credit portfolios, establishing an appropriate scaling factor could be a difficult task. In addition, there are important questions about whether the ability to validate model estimates would be meaningfully improved even using comparatively low percentiles of the loss distribution.[13]

> *As an alternative to a regulatory framework based on specific modeling restrictions,* conceptual *standards could be developed that would require banks subject to an internal-models capital requirement to develop and use a* comprehensive credit risk model.

Model Standards

Portfolio credit risk models would have to meet certain regulatory standards to be judged by supervisors as sufficiently comprehensive to be used for capital calculations. Given the current rapid state of evolution of these models, these standards should not be highly restrictive. That is, they should not require specific mathematical approaches or the use of particular "approved" models, since at present there is little basis for concluding that one specific approach to credit risk modeling is uniformly better than all others in all situations. Such requirements either would impede future modeling advances or would require frequent revision of regulatory standards to encompass innovations and advances in modeling.

As an alternative to a regulatory framework based on specific modeling restrictions, *conceptual* standards could be developed that would require banks subject to an internal-models capital requirement to develop and use a *comprehensive credit risk model*. Flexibility could be permitted in how the concepts are incorporated within any given model, subject to a supervisory review and approval process to ensure that the model was sufficiently comprehensive. Supervisors could work with the industry to develop sound-practice guidance, which could be used when assessing banks' models to make certain that models and assumptions fall within an acceptable range. This approach might result in a degree of disparity across banks; however, some disparities may be desirable if they reflect legitimate differences in how individual banks choose to model the risk factors that are most important to their business mix.[14] As long as banking supervisors can verify that a bank's choices are reasonable and that model parameters have a sound empirical basis, conceptual standards could strike a balance between ensuring comparability, on the one hand, and facilitating continued model improvement and innovation, on the other.

The rest of this section considers how modeling standards might address the conceptual elements that characterize comprehensive portfolio credit models as outlined earlier. The discussion covers the key elements of robust credit risk modeling to indicate a potential starting point for regulatory modeling standards. Conceptual standards for comprehensive models would have to cover two major areas: model structure and general data requirements related to parameter estimation and to the way in which portfolio structure is captured within the model.

Standards for Model Structure

Comprehensive credit risk models account for variation in and correlation between losses from individual credits, borrowers, or counterparties. This can be accomplished in a variety of ways, but in general terms it entails accounting for variation due to three key modeling elements: transition probabilities, credit exposures, and asset revaluation. Structural modeling standards would have to address all three areas.

Transition probabilities: In one way or another, comprehensive models incorporate the probability that any given position might have migrated to a different credit quality state at the planning horizon. In a default mode framework, this requires an assessment of the probability of default, while in a multistate framework, the model must capture the probabilities of credits moving from one credit state or risk category to any of the others. At a minimum, standards would require that models used for regulatory capital do this.

However, transitions between credit quality states are correlated to some extent across borrowers. Structural modeling standards would have to address the extent to which models should recognize this fact. A requirement that models incorporate this type of correlation should not pose a significant hurdle for most banks, because few if any models assume that variation in credit quality is independent across borrowers. This is hardly surprising, since a model that made such an assumption would fail to capture one of the most important influences on risk in a credit portfolio. A standard probably would also require that the relevant correlations be based on empirical analysis, although in some cases a more judgmental process might be warranted.

Credit exposures: Uncertainty in credit exposures at the horizon may stem from direct dependence on market prices or rates, such as counterparty credit risk exposures under derivatives contracts. It also may arise for other reasons, as in the case of lines of credit and standby letters of credit that depend on actions of borrowers that are generally beyond a bank's control. Because the size of credit exposures has a first-order effect on measured credit risk—for example, a 20 percent increase in exposure generally leads to a 20 percent increase in the risk estimate—standards for comprehensive models would have to specify an approach to recognizing this uncertainty.

At a minimum, a regulatory standard could require models to recognize that exposures can change, perhaps by making "stress case" assumptions about exposures at the end of the planning horizon. An example of such an approach would be to assume that all credit lines will be completely drawn down, or that derivatives will have exposures equal to some high percentile of their potential future values. In the near term, a realistic and adequate regulatory standard might simply require that models incorporate deterministic changes in exposures according to credit quality states, but a more complete alternative would be to incorporate an element of random variation in exposures.[15]

Comprehensive credit risk models [would] account for . . . variation due to three key modeling elements: transition probabilities, credit exposures, and asset revaluation.

For positions that involve derivatives or that otherwise depend to a material extent on market factors, standards likely would require integrated models of market movements and credit exposures. Especially in such cases, banks' credit risk models should reflect not only the uncertainty in future exposures, but also the potential correlation of exposures across credits. For example, a bank's counterparty exposures from derivatives contracts that are linked to a common market price will certainly be correlated, and this correlation should be captured in exposure estimates. This is an area in which modeling practice is developing rapidly, and fairly rigorous regulatory standards likely would be appropriate.

Asset revaluation: An integral part of any credit risk model is revaluing various credit exposures as they migrate across credit quality states. As noted in the prior section, in multistate models this process of asset valuation consists of revaluing positions according to their credit quality and the general market conditions expected at the end of the planning horizon, generally by using market credit spreads to discount contractual payments.

Standards for comprehensive models should require banks to capture not only the expected change in value as positions migrate across credit quality states, but also the impact of the uncertainty around these changes. Thus, using a market-based but fixed-term structure of credit spreads would be inadequate. Incorporating deterministic changes in credit spreads, perhaps based on the forward spreads implied in the yield curve, is more sophisticated but still does not capture the effects of uncertainty. Thus, modeling standards might require that volatility in market credit spreads and correlations between changes in these spreads be explicitly incorporated into revaluations due to migration across credit quality states.

Default states often are treated separately, with revaluation based on the fraction of the exposure that ultimately will be recovered. Recovery rates vary by facility type, across industries, and across countries. However, they also vary uncertainly with conditions in asset markets, and standards for comprehensive models probably would require banks to incorporate this source of uncertainty.[16] An important question in setting model standards is whether models should be required to capture correlations among recovery rates in addition to variation, and, if so, what sort of standards can reasonably be established to ensure that these correlations are adequately captured.

A comprehensive credit risk model must be based on a rating process that is sound and rigorous and that incorporates all relevant information, both public and proprietary.

Other aspects of correlation: As noted above, cross-credit correlations are important within each of the three dimensions of transition probabilities, exposures, and revaluation. However, there can also be important correlations across these dimensions. For example, the same factors that cause a borrower to transition to an inferior credit quality state might also cause an increase in the draw on a line of credit and a simultaneous decline in the value of collateral assets. In that case, all three dimensions of credit uncertainty are correlated.

Capturing these types of correlations is an area in which credit risk models have made limited progress. To date, most credit risk models assume that most of these correlations are zero. Model developers sometimes assert that such assumptions are appropriate because the correlations either are relatively unimportant or are impractical to model. Further exploration of such assertions would be necessary to ensure that these assumptions are reasonable. Standards for comprehensive models could require banks to either estimate and incorporate the relevant correlations or demonstrate convincingly that they are not material. This would likely present a significant hurdle, given the current state of model development.

Thus far, this section has outlined a qualitative standard requiring a model to capture correlations both within and across each of the three dimensions of transition probabilities, exposures, and revaluation. As noted earlier, nearly all models assume that these correlations are driven by one or more risk factors that represent various influences on the credit quality of the borrower. The assumptions about the statistical process driving these risk factors determine the overall mathematical structure of the model and the ultimate shape of the PDF. As such, a comprehensive models standard would need to address the underlying distribution of these risk factors.

Although it might be desirable to develop a specific standard for the distribution of the risk factors, differences in model structure again make it difficult to establish minimum requirements that would be broadly applicable. Given the importance of these embedded assumptions, the development of such standards may be one of the most important hurdles that banks and supervisors will need to clear before an IM approach for credit risk could be implemented. At a minimum, as an alternative, supervisors would need to address the calibration and statistical process driving these risk factors in sound-practice guidance.

Standards for Data and Estimation

Data requirements may pose some of the most significant implementation hurdles for an IM capital adequacy regime.[17]

Two major categories of data are required for models-based capital calculations. First, the credit portfolio must be characterized in some consistent way, appropriate to the model being used. That is, the portfolio structure must be captured. Second, any model relies on certain parameter estimates, typically computed from empirical observations, corresponding to the conceptual dimensions described above. These parameter estimates tailor the more general conceptual model of credit risk to the specific operating environment of a bank. This section discusses some general issues related to data, for both portfolio structure and parameter estimation, and the types of regulatory standards that might be appropriate for this aspect of credit risk modeling.

Portfolio structure: In a comprehensive credit risk model, the two most important aspects related to portfolio structure are that the portfolio be appropriately segregated by credit quality and that all material exposures be accounted for. The nearly universal approach within the industry for characterizing credit quality is to assign each exposure a numerical rating along a continuum of risk grades that divides the exposures into various categories according to credit risk. A number of different approaches are used in practice, based on some combination of external agency ratings, market and financial statement data, and other information. In marked contrast to market risk models, banks use internal analysis and private, proprietary information on relevant borrower and counterparty characteristics to determine how exposures are included in credit risk models. Sound practices in the area of internal credit risk rating have been evolving rapidly. Whatever approach a bank uses, the overall quality of the credit risk modeling effort depends heavily on the quality of the rating process. Thus, a comprehensive credit risk model must be based on a rating process that is sound and rigorous and that incorporates all relevant information, both public and proprietary. Standards in this area are the subject of ongoing efforts by regulatory and industry groups.

Aside from being based on a rigorous credit rating system, a comprehensive credit risk model must capture all material credit exposures and incorporate them appropriately in the calculations. This process would start with identifying which positions within a bank's portfolio were subject to the credit risk capital charges. The current regulatory capital structure separates positions into those subject to market risk capital standards and those subject to credit risk standards, primarily on the basis of whether a position is held inside or outside of a bank's trading account. Thus, a clear delineation between the banking and trading books would be necessary to prevent "regulatory arbitrage" intended to minimize regulatory capital requirements by inappropriately shifting positions across books. Of course, such incentives exist even in the absence of an IM approach to credit risk, and supervisors have developed guidance to govern the treatment of various types of positions. To the extent that the incentives to engage in such regulatory arbitrage are heightened under an IM regime, supervisors could refine this guidance to ensure that it limits the opportunity for banks to shift positions solely to benefit from reduced capital requirements.

Once the positions subject to the credit risk capital requirements have been identified, regulatory standards would require institutions to demonstrate that their information systems consolidate credit exposure data globally, with any omissions immaterial to the overall credit risk profile of the institution. For completeness, the structural data would have to capture the flow of new credits into each rating category, the elimination of any retiring credits, and the migration of existing credits into other rating categories. That is, initial ratings should be updated periodically to reflect the current financial condition of borrowers or counterparties. In addition, the model should aggregate all material exposures for each borrower, so that a consolidated exposure estimate is produced.

Banks would be expected to explain and justify estimation methods to bank supervisors and to provide sufficient support—such as literature citations, technical documents, and access to developers—to make possible a rigorous assessment of the parameter estimation methodology.

Parameter estimates: Parameter estimation gives rise to some of the most significant data issues in constructing a comprehensive credit risk model. Estimation techniques often are unique to a particular model, so again the standards must be conceptual rather than specific. However, banks would be expected to explain and justify estimation methods to bank supervisors and to provide sufficient support—such as literature citations, technical documents, and access to developers—to make possible a rigorous assessment of the parameter estimation methodology.

Data sources vary by type of parameter. Data on transition probabilities may come from a bank's own credit migration experience. In contrast, parameters that reflect state values and their variations generally are based on market credit spread data, estimated from historically realized values on asset sales for certain types of assets, or based on recovery rates for assets in default. Whatever the specific data used to calibrate the parameters, regulatory standards likely would reflect three general principles. First, the data should be drawn from a historical period that reflects a wide range of potential variation in factors related to credit quality, thereby providing adequate *historical coverage*. Second, the data should be *applicable to the specific business mix* of the bank. Third, the data should reflect consistent *definitions of default* or of relevant credit-state transitions.

With regard to historical coverage, a comprehensive approach would require that the data, in combination with the model structure, be sufficient to reflect credit cycle effects. To achieve that, regulatory standards likely would require a historical window that encompasses a period sufficiently long to capture defaults and downgrades that were at various times both high and low by historical standards. Specific requirements may vary depending on the asset type, geographic region, or product market in question, since different products and markets experience cycles at different times and with different frequencies, but an adequate window would almost always span many years.

> *The supervisory validation process can be viewed as comprising the following two elements. The first is the development of sound-practice guidance for the structure and implementation of credit risk management models. . . . The second element . . . is the use of quantitative testing to detect systematic biases in model results.*

With regard to bank-specific applicability, regulators probably would expect a bank to be able to demonstrate that the data used to estimate model parameters are appropriate for the current composition of its portfolio. For example, data from U.S. corporations might not be appropriate for use in models that cover exposures to European or Latin American borrowers. Similarly, transition probabilities or state-valuation estimates based on national level data might be inappropriate for institutions with loan portfolios that contain highly specific regional or industrial concentrations.

At least in the near term, banks and supervisors are likely to face a trade-off between the dual requirements of data applicability and coverage of the historical window. Using a bank's own internal data generally solves the applicability problem, as long as any significant historical changes in the bank's business profile are addressed and provided the bank has experienced a sufficient number of defaults and losses to produce reasonably accurate parameter estimates. However, at present it appears that few banks can construct an adequate data history based on internal data. Alternatively, banks could use vendor-provided or public data—for example, data from publicly traded bonds—or pooled data from a group of peer institutions to estimate parameters. Since historical data of this type are more readily available, issues related to sample period and coverage of the credit cycle can be addressed more easily, but demonstrating that the results are applicable to a specific bank's business mix becomes more difficult.

Finally, parameter estimates should be based on common definitions of default or, in a multistate framework, common definitions of credit-state transitions. Inconsistency in the data used could lead to highly erroneous estimates. It may be particularly important to ensure that the data used for default probabilities and associated losses-given-default reflect consistent definitions. For example, if default probabilities calculated from publicly traded bond data were combined with loss-given-default figures from internal bank data on nonaccrual loans, the resulting estimates of risk could be seriously understated, owing to the less severe credit events defined as "default" in the internal data. This type of definitional issue also may be especially problematic when data are drawn from multiple bankruptcy regimes, as is generally the case for international data.

Validation

The third component of an IM capital regime concerns supervisory model validation, that is, the process of ensuring that the model is implemented in a rigorous way.[18] As in the discussion of the structure of an IM capital regime for credit risk, it is useful to begin this discussion by recalling the validation approaches applied in the market risk setting. The market risk validation approach relies on a combination of qualitative standards and statistical testing. The qualitative standards address the internal controls and procedures surrounding the design and operation of the models used for regulatory capital purposes, focusing on issues such as the need for an independent risk management function, regular risk reporting to senior management, and periodic independent audits of the model. In addition to the qualitative standards, supervisory validation also involves statistical testing of the output of the market risk measurement models, or so-called back-testing. Back-testing is a way of assessing the accuracy of a model's estimate of the target percentile of the loss distribution—the 99th percentile in the case of the market risk capital charge—through a comparison of the actual gains and losses on the trading portfolio with the risk estimates supplied by the model.

Against this background, the supervisory validation process can be viewed as comprising the following two elements. The first is the development of sound-practice guidance for the structure and implementation of credit risk management models. This guidance would consist of a largely qualitative description of the current state of the practice in credit risk measurement, covering both technical aspects of model design and estimation and qualitative standards for the risk management environment. The technical aspects of model design would cover the elements of a comprehensive credit risk model, as indicated above, while the qualitative standards would focus on the policies and procedures used by the bank in its risk management activities. A key element among these policies and procedures would be a "use test" to ensure that any model used for regulatory capital purposes is in fact an integral part of the bank's risk management structure.

The second element of the supervisory validation process is the use of quantitative testing to detect systematic biases in model results. Unlike in the market risk setting, formal back-testing of credit risk model results is not feasible because of the length of a typical credit cycle and the resultant limited number of independent observations of actual outcomes.[19] As a result, model validation in the credit risk setting will likely have to draw on a combination of tests, at least until more internal data become available and more robust statistical methodologies are developed. These tests could consist of both work that banks have done internally as part of model design and upkeep (for example, sensitivity tests of key parameters) and supervisory tests intended to identify systematic differences across banks in model outputs (for example, "test portfolio" exercises). Finally, public disclosures about model design, estimation, and output are another way to bring scrutiny to the models used by banks for capital purposes. All of these elements together are intended to provide both supervisors and the banks themselves assurance that any model used for regulatory capital purposes is theoretically sound and properly implemented.

Sound-Practice Guidance

The purpose of sound-practice guidance would be to describe in more detail the various elements that supervisors would consider when evaluating internal models used for regulatory minimum capital calculations. In some cases, guidance would describe standards that any model should meet to be considered accurate. In other cases, guidance would serve to reflect the range of practice found at banks with more advanced approaches to modeling. Guidance on sound practices would be dynamic and change over time to reflect the then-current state of the practice in credit risk modeling, providing both supervisors and banks with a benchmark against which to assess a particular model's structure and implementation. In particular, since credit models will almost inevitably incorporate a certain degree of management judgment—for instance, simplifying assumptions about correlations or other parameters, the use of less-than-perfect data to calibrate model parameters, or assumptions about the distribution of aggregate losses—the guidance could provide a way of assessing these assumptions against the range of current practice.

The experience gained by both banks and supervisors in implementing the revised Basel Accord has the potential to provide important insight into the development of qualitative standards and for validation more generally.

Within the supervisory process, there is growing emphasis on qualitative reviews of banks' methods for measuring, managing, and controlling their risk exposure and the implications for capital adequacy.[20] A key part of any sound-practice guidance would be qualitative standards for the risk management environment. Supervisors have developed significant experience using qualitative sound practice standards to assess banks' risk management processes in the context of market risk. Finally, the upcoming revisions to the Basel Accord will likely incorporate a greater reliance on banks' internal risk rating systems in assessing regulatory minimum capital requirements. The experience gained by both banks and supervisors in implementing the revised Basel Accord has the potential to provide important insight into the development of qualitative standards and for validation more generally.

Building on the precedent of the market risk amendment to the Basel Accord, banks' use of portfolio credit models for regulatory capital purposes would be contingent upon their meeting a series of qualitative standards aimed at ensuring that the models used are sound and implemented with integrity. Qualitative standards aimed at aligning banks' risk management techniques with supervisory safety and soundness objectives could include:

- compliance with a documented set of internal policies, controls, and procedures concerning the operation of the credit risk measurement system
- an independent risk control unit responsible for the design and implementation of the bank's credit risk model
- a regular independent review of the credit risk model as part of the bank's own internal auditing process, either by internal or by external auditors.

Finally, the qualitative guidelines should incorporate a "use test" to ensure that any model used for regulatory capital is closely integrated with the ongoing credit risk management process of the bank. In particular, the model's output should be an integral part of the process of planning, monitoring, and controlling the bank's credit risk profile. For instance, the model might be used in conjunction with internal credit exposure limits, capital and portfolio allocation decisions, or pricing. All of these uses suggest that a bank would have significant incentives to invest sufficient resources in model development and maintenance to ensure that the model is producing reliable risk estimates. Just as in the IM approach for market risk, where a use test is one of the key elements of the qualitative guidelines, such tests could help to provide discipline in the credit risk setting.

Quantitative Testing

Aside from ensuring that a model meets sound-practice standards, supervisory validation could include empirical testing of the model's inputs and results. Given the shortcomings of formal back-testing of model results, quantitative testing in the credit risk setting is likely to rest on independent review of model parameters such as expected default probabilities, loss given default, and exposure estimates; sensitivity analysis of key parameter estimates; stress testing of model results; and test portfolio exercises intended to identify possible systematic biases in model outcomes across banks.[21]

These elements provide a possible roadmap for quantitative testing for model validation, but considerable work would be required to implement these ideas in an effective way. With the exception of the test portfolio exercises, this quantitative testing would most likely build on the work done by the banks themselves as part of their internal-model development and maintenance procedures. That is, the first step in a supervisory review of a bank's credit risk model should be the review of the bank's own work papers documenting the tests done by the model builders and by the bank's internal or external auditors to calibrate and test the model.

To support this process, supervisors could develop sound-practice guidance on the types of tests that banks would be expected to perform as part of developing and maintaining their credit risk models. For instance, testing could include sensitivity analysis—that is, analysis of the sensitivity of the model results to changes in parameters and key assumptions. Sensitivity analysis allows management to probe the vulnerabilities in a model that arise from its structure, use of a particular type of statistical technique, or limitations in terms of historical observations. This analysis might include demonstrating the impact on the model's output and resulting capital charge from changes in recovery rates, correlations, and credit spreads. In all cases, banks would likely be expected to maintain adequate documentation to permit a rigorous review of model development and testing.

This list is not intended to be exhaustive, because these tests would by neccessity be somewhat model-specific. However, there are likely to be a general range of parameters and assumptions that banks could be expected to examine. Where the analysis indicates that particular parameters and assumptions have a significant impact on the model results, the sensitivity analysis should yield a thorough understanding of the impact of changes.

While stress testing is far from a perfect validation tool, it can provide important information about the impact of unlikely but potentially damaging events that could result in very large losses in a bank's credit portfolio.

Another important benchmark against which supervisors can assess the reasonableness of a bank's modeled capital requirement is stress testing. Stress testing is an important element of the modeling and risk management process that can help ensure that potentially large portfolio losses are not hidden by overly optimistic or simplistic assumptions. While stress testing is far from a perfect validation tool, it can provide important information about the impact of unlikely but potentially damaging events that could result in very large losses in a bank's credit portfolio.

The key, of course, is identifying a meaningful range of stress scenarios and accurately assessing their likely impact on the credit portfolio.[22] These stress scenarios could involve actual historical events; simulated increases or decreases in the model's transition probabilities, volatilities, or correlations; or a widespread deterioration in credit quality. As challenging as identifying meaningful stress scenarios might be, the lack of historical data and the inability to back-test model results make stress testing an important and independent indicator supervisors can use for gauging the reliability of the modeling process and the appropriateness of the resulting capital charge. As such, it is an important tool in the arsenal for the evaluation of credit risk models.

Beyond the testing done by the bank, supervisors may want additional verification that the model output is reasonable. However, the absence of back-testing requires that supervisors rely on other tools to help them evaluate the output of a bank's credit risk measurement model and to serve as a foundation for dialogue and discussion with the bank. Possible tools include supervisory stress tests, the use of test portfolios, and supervisory use of vendor-provided models.

Public Disclosure of Model Specifications

Another approach to model validation, somewhat different from the supervisory processes described above but possibly complementary to them, would be to require all banks using internal models for regulatory capital purposes to disclose publicly full documentation of the model's mathematical structure, key assumptions, and parameter estimates.[23] The purpose of such disclosure would be to expose the bank's model to the discipline of public scrutiny. This scrutiny could aid the supervisory validation process by providing independent assessments of a bank's model by market practitioners and interested academics. In addition, it could improve modeling practices for the industry as a whole by ensuring that the latest modeling innovations were quickly disseminated to all practitioners.

While the benefits of such disclosure could be substantial, they would depend on the ability of supervisors and banking institutions to establish a workable disclosure framework. In principle, this could be accomplished through regulatory disclosure requirements, though these could be difficult to define in view of the wide variety of models and the continuing rapid evolution in industry practice. Alternatively, disclosures could be assessed through the supervisory review process to ensure that key elements of model structure and design were being accurately portrayed. The benefits of public disclosure would also have to be weighed against their potential costs, including the possibility that mandatory disclosure would undercut banks' incentives to develop new and innovative modeling practices, since they would have to share the benefits of any innovations with their competitors.

Summary and Conclusions

In this article, we have attempted to lay out the issues that would have to be addressed in creating a regulatory minimum capital requirement based on the output of banks' internal credit risk models. Using the current market risk capital requirements as a guide, we identified three basic components of an IM credit risk capital charge: prudential standards defining the risk measure to be used in the requirement, modeling standards describing the essential components of a comprehensive credit risk model, and validation standards governing the techniques used by banks and by supervisors to ensure that the models are conceptually sound and reasonably accurate. An important consideration in specifying standards in these three areas would be to balance the desire for flexibility and innovation in modeling practice, on the one hand, with the need to ensure that the capital charge is conceptually sound, empirically accurate, and reasonably comparable across banks, on the other.

This article is not intended to be a policy proposal. Instead, our goal is to stimulate discussion among financial institutions, supervisors, and other interested parties about the many practical and conceptual issues that would be involved in structuring a workable IM regulatory capital regime for credit risk. The Basel Committee is in the process of revising regulatory capital standards, and a key factor in considering any IM regulatory capital regime will be the experience of both supervisors and financial institutions with these new, more risk-sensitive standards. For these reasons, the discussion above should be interpreted as an initial step in trying to establish some general principles that could guide the ultimate formation of an IM approach to regulatory capital rather than any kind of definitive statement of what such an approach would look like.

As our discussion suggests, the challenges in developing an IM framework would be significant, both for banks and for supervisors. These challenges involve the further technical development of the credit risk models used by financial institutions, the accumulation of improved data sources for model calibration, and the refinement of procedures used by banks and supervisors to validate the accuracy of the models' risk estimates. In addition, a variety of detailed implementation issues would have to be worked out (see the appendix for a discussion of these points). Our hope is that this article will represent a constructive step in identifying the most important of these many issues and in helping to determine the feasibility of an IM approach for credit risk.

Appendix: Practical Implementation Issues

A number of issues not discussed in this article would have to be addressed before an internal-models-based (IM) approach to regulatory capital for credit risk could be implemented. These issues include:

- *Loan loss reserves and expected loss*: A capital charge based on unexpected losses raises important issues concerning the role and definition of loan loss reserves. Recall that unexpected losses equal losses at the target percentile minus expected losses. Therefore, if loan loss reserves fall short of expected losses, the total resources available to absorb losses—reserves plus capital—will not be sufficient to provide protection at the desired soundness standard. Unfortunately, there is no necessary correspondence between the accounting definition of loan loss reserves and the concept of expected losses from a credit risk measurement model. Thus, over the longer run, basing a regulatory capital charge on unexpected losses may require a rethinking of the treatment of loan loss reserves.
- *Eligible institutions*: The set of institutions subject to an IM capital requirement will most likely be defined by the minimum standards that are developed. Initially, only a small set of banks would likely have models that were sufficiently well-developed; many banks currently employ default mode models and few, if any, fully capture the correlation between risk drivers such as the potential correlation between defaults and recovery rates. Over time, however, the set of institutions with comprehensive credit risk models is likely to grow as modeling expertise disseminates through the industry, as data sources become more readily available, and as the competitive incentives for institutions to manage their credit risk exposures in a more active way intensify.
- *Scope of application*: An important issue is whether an IM capital requirement could be designed to cover all of a bank's credit exposures, or only those in selected portfolios (for instance, large commercial loans). The models discussed in this article are applied primarily to commercial lending portfolios, while other portfolios—such as retail lending—are either covered by models whose structures are very different or, occasionally, not covered at all. In this situation, it might make sense to allow banks to apply an IM capital requirement only to those portfolios covered by comprehensive credit risk models of the type described here and to use a non-models-based regulatory capital requirement for other portfolios. However, "cherry picking," or selective adoption, is a clear concern if banks are allowed to use internal models to determine capital charges for some, but not all, of their exposures. That is, a bank may have an incentive to model only those portions of its portfolio in which capital charges are reduced.
- *Scaling factor*: The IM capital requirement for market risk incorporates a multiplicative scaling factor that is intended to translate value-at-risk estimates into an appropriate minimum capital requirement, reflecting considerations both about the accuracy of a bank's value-at-risk model and about prudent capital coverage. There could be a similar role for a scaling factor in an IM credit risk capital regime. For instance, given shortcomings in data availability, uncertainty surrounding the calibration of credit risk model parameters (so-called model uncertainty) is a significant concern in using these models for regulatory capital purposes. More generally, supervisors and banks lack long-term experience with credit risk models, a fact that creates uncertainty about how the models will perform over future credit cycles and during times of financial market distress. These concerns could be addressed—albeit roughly—by scaling up the raw loss figures reported by the banks. In this instance, a scaling factor might be incorporated when an IM approach is initially implemented, and then revisited as both supervisors and banks gain experience with the IM regime.
- *Frequency of capital calculations*: Prudential standards would have to specify how frequently banks would be required to run their credit risk models and report the results to supervisors. Unlike value-at-risk models, which are run on a daily basis to assess the market risk in banks' trading activities, credit risk models are run less frequently. Monthly runs of the model—where a "run" of the model means a new estimate of the PDF of future losses incorporating changes in portfolio composition, credit ratings, market prices, and parameter updates, where warranted—seem a reasonable minimum standard in the near term, though over the longer run, banks would probably be expected to develop the capability to generate fresh model estimates on an even more frequent basis (perhaps weekly or biweekly).

Given frequent model results, capital could be based on the average of monthly or weekly estimates during the quarter. Using an average should mitigate banks' incentives to window dress, as might be the case if the capital charge were based on model outputs as of a single point in time, such as quarter-end. In addition, averaging should smooth short-run volatility in the model estimates and ensure that the capital requirement is not overly sensitive to short-term anomalies in credit markets.

- *Parameter updates*: A bank using an internal model for credit risk capital would also be required to update the model's inputs and parameters with some minimum frequency. There are obvious trade-offs between accuracy of risk assessment and reporting burden: more frequent updating gives regulators and banks more confidence in the model results, but may impose a greater burden on the banks. Different updating schedules may be reasonable for different types of parameters and different data sources. For instance, many models use market-based credit spreads to revalue credit exposures. These spreads should be updated frequently, probably more frequently than the full model is reestimated, to account for the potentially significant variation of spreads over relatively short periods. In contrast, state-value estimates based on recovery rates or on market prices from asset sales could be updated less frequently, as could transition probabilities and correlations, although additional work would be desirable to confirm the optimal timing. Portfolio structure data should be updated at least as often as the material is run.

Endnotes

1. This section draws heavily on a recent Federal Reserve study of the structure and implementation of credit risk models at large U.S. banking institution (Board of Governors of the Federal Reserve System 1998b). For interested readers, this paper contains an in-depth discussion of credit risk modeling issues.

2. A discussion of internal risk rating systems is beyond the scope of this article. However, since sound-practice standards and guidelines for internal rating systems are under active consideration as part of the Basel Committee's efforts to revise capital standards, regulators' expectations regarding such rating systems will become known as part of that process. For further discussion regarding internal rating system standards, see Board of Governors of the Federal Reserve System (1998a). In addition, for information on the range of internal rating practices among international banks, see Treacy and Carey (1998) and Basel Committee on Banking Supervision (2000).

3. The LGD, sometimes also referred to as the loss in event of default, is equal to 1 minus the recovery rate on the defaulted loan.

4. See, for example, Gordy (2000a).

5. See Basel Committee on Banking Supervision (1996) and Hendricks and Hirtle (1997) for a full description and discussion of the market risk capital requirements.

6. It is interesting to note that under a lifetime horizon, there is no distinction between the multistate and default-mode loss definitions since credits will either default or mature over their lifetimes; intermediate upgrades and downgrades short of default have no impact on the value of credits at the horizon.

7. One concern that arises in specifying a given planning horizon for regulatory capital purposes is that this choice may impede supervisors from urging banks to use different planning horizons for internal purposes if market best practice evolves over time. In the market risk setting, this concern is addressed through a simple scaling approach, where capital requirements based on a ten-day standard may be calculated with scaled risk estimates based on the one-day horizon that is typical for most value-at-risk models. However, the nature of the processes underlying credit risk is sufficiently different that this approach may not be acceptable. For credit risk, it may be more appropriate for supervisors to address such issues through the review of banks' internal capital allocation methodologies (see Board of Governors of the Federal Reserve System [1999]). Another alternative would be to allow each bank to use a bank-specific planning horizon—or even a bank-specific loss definition/planning horizon/ target loss percentile combination—but this approach would introduce very significant problems in establishing a consistent minimum regulatory capital requirement across banks.

8. Subtracting the expected loss from the specified loss percentile reflects the concept that capital is used primarily to cover unexpected losses. Regulatory standards would have to ensure that expected losses were covered in other ways, such as through loan loss reserves or through credit spreads on the pricing of credit extensions. See the appendix for a more detailed discussion.

9. As an alternative to value at risk, some have suggested that using the "expected tail loss"—that is, the expected loss given that the loss is greater than the target percentile—as a measure of risk. See, for instance, Gordy (2000b).

10. For instance, the historical insolvency rate on AA-rated bonds is about 0.03 percent, implying that a target percentile of 99.97 would be required to provide that degree of coverage. A 99.97th target percentile would mean that unexpected losses would exceed capital only 0.03 percent of the time.

11. See Estrella (1995) and Basel Committee on Banking Supervision (1999b) for a discussion of the role of minimum capital requirements in regulation and supervision.

12. Establishing higher regulatory capital requirements than banks themselves would select on safety and soundness grounds would imply that supervisors were having an inappropriate impact on banks' business decisions. Under the current capital standard, this phenomenon sometimes encourages banks to securitize assets when regulatory capital requirements exceed what the market demands.

13. For instance, even using a 90th percentile figure, we would expect to see losses exceeding this level only once every ten years. Further, validation procedures that examine the entire tail of the distribution—rather than a single point at a given percentile—may prove more powerful in identifying models that fail to capture extreme loss behavior. In that event, the ability to validate models would depend much less on the particular percentile chosen to form the basis of the capital requirement.

14. A study by the Institute of International Finance, Inc., and the International Swaps and Derivatives Association (2000) highlights not only some of the differences that can result from banks' different

modeling choices, but also differences that arise from the calibration and implementation of models that are otherwise similar.

15. Asarnow and Marker (1995) present an empirical study of the relationship between the use of lines of credit in the event of default and borrower credit quality.

16. Frye (2000) highlights the challenges and potential importance of incorporating recovery rate uncertainty.

17. See Basel Committee on Banking Supervision (1999a) for a full discussion of these hurdles.

18. This section focuses on validation of portfolio credit models. Critical validation issues also arise with regard to the mappings of individual credits into a given institution's internal credit grades. As indicated above, both banks and supervisors have been devoting significant attention to this process in recent years, and considerable progress has been made in addressing some of the key issues. See, for instance, Basel Committee on Banking Supervision (2000) and Board of Governors of the Federal Reserve System (1998a).

19. Lopez and Saidenberg (2000) discuss some of the challenges and limitations of back-testing credit risk models.

20. See Board of Governors of the Federal Reserve System (1999) for a discussion of banks' internal capital allocation procedures and Basel Committee on Banking Supervision (1999b) for a discussion of internal capital as "pillar two" of the proposed revisions to the Basel Accord.

21. In a test portfolio exercise, supervisors construct one or more standard portfolios, which may be composed of actual or hypothetical credit positions, and each bank is asked to produce risk estimates for these portfolios using its internal model. The resulting figures are then compared across institutions to generate a sense of the range of model outcomes and potentially to identify "outliers" whose risk estimates fall outside the typical range.

22. Berkowitz (2000) discusses the challenges of establishing a comprehensive approach to stress testing.

23. Basel Committee on Banking Supervision (1999b) and Estrella (1995), among others, outline the importance of disclosure and market discipline as components of banking regulation and supervision.

References

Asarnow, Elliot, and James Marker. 1995. "Historical Performance of the U.S. Corporate Loan Market: 1988-1993." Commercial Lending Review 10, no. 2: 13-32.

Basel Committee on Banking Supervision. 1996. "Amendment to the Capital Accord to Incorporate Market Risks." Basel, Switzerland: Bank for International Settlements.

———. 1999a. "Credit Risk Modeling: Current Practices and Applications." April. Basel, Switzerland: Bank for International Settlements.

———. 1999b. "A New Capital Adequacy Framework." June. Basel, Switzerland: Bank for International Settlements.

———. 2000. "Range of Practice in Banks' Internal Ratings Systems." Discussion Paper no. 66. Basel, Switzerland: Bank for International Settlements.

Berkowitz, Jeremy. 2000. "A Coherent Framework for Stress-Testing." Board of Governors of the Federal Reserve System Finance and Economics Discussion Series no. 1999-29.

Board of Governors of the Federal Reserve System. 1998a. "Sound Credit Risk Management and the Use of Internal Credit Risk Ratings at Large Banking Organizations." SR Letter 98-25. September 21.

———. 1998b. "Credit Risk Models at Major U.S. Banking Institutions: Current State of the Art and Implications for the Assessments of Capital Adequacy." May.

———. 1999. "Assessing Capital Adequacy in Relation to Risk at Large Banking Organizations and Others with Complex Risk Profiles." SR Letter 99-18. July 1.

Estrella, Arturo. 1995. "A Prolegomenon to Future Capital Requirements." Federal Reserve Bank of New York Economic Policy Review 1, no. 2: 1-12.

Frye, Jon. 2000. "Collateral Damage." Risk Magazine 13, no. 4: 91-4.

Gordy, Michael. 2000a. "A Comparative Anatomy of Credit Risk Models." Journal of Banking and Finance 24, nos. 1-2 (January): 119-49.

———. 2000b."A Risk-Factor Model Foundation for Ratings-Based Bank Capital Rules." Unpublished paper, Board of Governors of the Federal Reserve System. September.

Hendricks, Darryll, and Beverly Hirtle. 1997. "Bank Capital Requirements for Market Risk: The Internal Models Approach." Federal Reserve Bank of New York Economic Policy Review 3, no. 4 (December): 1-12.

Institute of International Finance, Inc., and International Swaps and Derivatives Association. 2000. "Modeling Credit Risk: Joint IIF/ ISDA Testing Program."

Lopez, Jose A., and Marc R. Saidenberg. 2000. "Evaluating Credit Risk Models." Journal of Banking and Finance 24, nos. 1-2 (January): 151-65.

Treacy, William F., and Mark Carey. 1998. "Credit Risk Rating at Large U.S. Banks." Federal Reserve Bulletin 84: 897-921.

Kevin J. Stiroh

What Drives Productivity Growth?

- Neoclassical and "new growth" theories offer alternative explanations for productivity and output growth.

- In the neoclassical view, exogenous technical progress drives long-run productivity growth since broadly defined capital suffers from diminishing returns. In contrast, the new growth models yield long-run growth endogenously, either by avoiding diminishing returns to capital or by explaining technical progress internally.

- Despite their differences, both views help to explain the recent rise in U.S. productivity growth. The methodological tools developed by neoclassical economists provide a means to measure the rate of technical progress, while the models of the new growth economists can provide an internal explanation for technical progress.

In 1995, the U.S. economy started to experience a strong resurgence in labor productivity growth. After growing only 1.3 percent per year from 1973 to 1995, labor productivity growth jumped to 2.5 percent from 1995 to 1999 (see chart).[1] This striking revival has hardly gone unnoticed, with academics, policymakers, and the financial press hotly debating competing explanations. Some commentators emphasize rapid capital accumulation and the recent investment boom, others point to deeper factors like fundamental technological change in high-tech industries, and still others argue that cyclical forces provide the primary explanation.[2]

This debate about the forces driving the U.S. economy mirrors a larger debate between the neoclassical and new growth theories regarding the sources of economic growth. Economists have long disagreed about this vital question, and the recent U.S. productivity revival presents an opportune backdrop to review this debate.

In the neoclassical view, broadly defined capital accumulation drives growth in the short run, but capital eventually succumbs to diminishing returns, so long-run productivity growth is entirely due to exogenous technical progress. The new growth theory, however, moves beyond this unsatisfying conclusion, arguing that productivity growth can continue indefinitely without the elixir of exogenous, and entirely unexplained, technical progress. Either by avoiding

Kevin J. Stiroh is a senior economist at the Federal Reserve Bank of New York.

An earlier version of this article was written for *Industry Canada* while the author was part of the Program on Technology and Economic Policy, Harvard University, Cambridge, Massachusetts. The author thanks Mun Ho, Dale Jorgenson, Kenneth Kuttner, Frank Lee, Charles Steindel, Lauren Stiroh, and two anonymous referees for helpful comments on an earlier draft, Mike Fort and Amir Sufi for excellent research assistance, and *Industry Canada* for financial support while the author was in the Program on Technology and Economic Policy. The views expressed are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

diminishing returns to capital or by explaining technical change internally, this framework offers an economic explanation for sustained productivity and output growth.

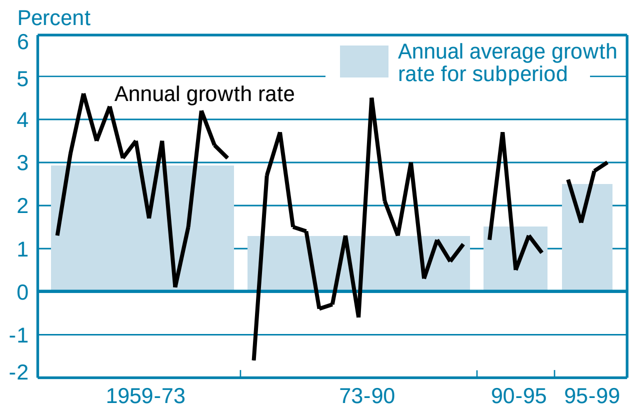
Despite these divergent conclusions, the neoclassical and new growth frameworks both contribute to our understanding of the growth process. Using traditional neoclassical methods, for example, Jorgenson and Stiroh (2000) and Oliner and Sichel (2000) show that a combination of accelerating technical progress in high-tech industries and the resultant investment in information technology (IT) are driving recent productivity gains in the United States. This type of neoclassical analysis clearly illustrates *what* happened to the U.S. economy. It cannot, however, explain *why* technical progress accelerated in high-tech industries; this is a job left to the new growth theorists. In this sense, each theory makes a significant contribution to our understanding of productivity growth. The sophisticated methodological tools developed by neoclassical economists enable us to measure the rate of technical change, while the sophisticated models of the new growth theorists provide an internal explanation for the sources of technical change. In the next section, I compare these alternative views and discuss the useful role of each.

One important theme, common to both views, is that investment is a fundamental part of the growth process. Investment, moreover, may be defined broadly to include any expenditure that provides productive payoffs in the future; therefore, measures of human capital and research and development (R&D) expenditures are now routinely included in productivity analyses. Indeed, even the concept of tangible capital is not static—the U.S. national accounts now treat software expenditures as an investment good since software code is a durable asset that contributes to production over several years.

In addition to this broadening of the investment concept, an important part of the measurement process is the recognition of the enormous heterogeneity of investment. For example, the productive impact of investment in information technology may be quite different from that of investment in structures. By disaggregating investment and accounting for these differences, economists can accurately gauge the productive impact of input accumulation and thus isolate and gauge the extent of technical change. In this article, I outline several major conceptual and methodological issues related to measuring production inputs and technical progress correctly.

The differences between neoclassical and new growth theories also have a direct bearing on several specific topics that have recently generated considerable interest. Later in this article, I address two relevant issues. In particular, I present a resolution to the computer productivity paradox and review the renewed embodiment controversy. These topics are important in their own right, and they illuminate the ongoing debate over the sources of productivity and output growth.

U.S. Labor Productivity Growth, 1959-99



Source: U.S. Department of Labor, Bureau of Labor Statistics.

Note: Both series refer to annual nonfarm business productivity.

Productivity Growth from the Neoclassical and New Growth Perspectives

Economic growth theory has enjoyed a revival in recent years, with questions about the sources of productivity growth high on the list of scholars, policymakers, and the business press. The academic growth literature has bifurcated, however, into arguments for two competing views—a neoclassical and a new growth view.[3] This section presents stylized models from each perspective and discusses the strengths and weaknesses of each.

Broadly defined investment—which includes expenditures on tangible assets, education, training, and other human capital accumulation, as well as research and development—plays a pivotal role in both frameworks, although investment's precise impact on productivity growth differs. Benefits from investment may accrue only internally to the economic agent that actually makes the investment, or the benefits may spill over more broadly to others in the economy. As a preview of the discussion, the idea that broadly defined capital generates primarily internal and diminishing returns is the hallmark of the neoclassical view, differentiating it from the new growth theory's focus on external and constant (or increasing) returns. This leads to contrasting views of the investment-productivity nexus and the potential for long-run growth.

The World of Neoclassical Growth

The standard neoclassical growth model is well known and will be reviewed here only briefly. Seminal papers by Solow (1956, 1957) formalized the neoclassical model, integrated theory with national account data, and formed the basis for much of applied growth analysis.

The Basic Neoclassical Model

The link between output, Y, and capital services, K, labor input, L, and labor-augmenting (Harrod-neutral) technical progress, T, is given by the familiar aggregate production function

(1) $$Y_t = f(K_t, T_t \cdot L_t),$$

where the neoclassical production function is typically assumed to have constant returns to scale, positive and diminishing returns with respect to each input, and marginal products of each input that approach zero (infinity) as each input goes to infinity (zero).

Investment enters through the capital accumulation equation, which governs the relationship between investment in tangible assets, I, and capital stock, S, via the perpetual inventory relationship

(2) $$S_t = (1-\delta) \cdot S_{t-1} + I_t,$$

where δ is depreciation and I_t can either be determined endogenously by profit-maximizing firms or assumed to be some fixed proportion of output.

Note that the production function includes a measure of capital services, K, while the perpetual inventory equation defines the capital stock, S. These two capital concepts are closely linked but differ according to compositional changes in aggregate investment. In the simplest neoclassical world with one investment good, the two concepts are identical, but in a world with many heterogeneous types of investment goods, they differ. For example, a shift in investment toward high-tech equipment with large marginal products leads capital services to grow more quickly than capital stock. This important distinction is discussed in detail later in this article.

The striking implication of the neoclassical model is that, in the long run, per capita output and productivity growth are driven entirely by growth in exogenous technical progress and they are independent of other structural parameters like the savings rate.[4] If the savings rate and investment share increase, for example, the long-run level of productivity rises but the long-run growth rate eventually reflects only technical progress. In this sense, the neoclassical growth model is not really a model of long-run growth at all since productivity growth is due to exogenous and entirely unexplained technical progress. Nonetheless, the neoclassical model has proved to be a useful tool for understanding the proximate factors that contribute to output and productivity growth.[5]

Solow (1957) provides an explicit methodology for measuring the rate of technical progress under the neoclassical assumptions of competitive factor markets and input exhaustion when technology is Hicks-neutral and output is modeled as $Y_t = A_t \cdot f(K_t, L_t)$. In this case, the rate of Hicks-neutral technical progress equals the famous Solow residual, or total factor productivity (TFP) growth. This is defined as the difference between output growth and the share-weighted growth rates of primary inputs (capital and labor) as

(3) $$\Delta lnA = \Delta lnY - v_K \Delta lnK - v_L \Delta lnL,$$

where Δ represents a first difference, v_K is capital's share of national income, v_L is labor's share of national income, the standard neoclassical assumptions imply $v_K + v_L = 1$, and time subscripts are dropped for ease of exposition.[6]

Under the same assumptions, one can identify the sources of average labor productivity (ALP) growth, defined as output per hour worked, Y/H. Transforming equation 3 yields

(4) $$\begin{aligned} \Delta lny &= \Delta lnY - \Delta lnH \\ &= v_K \Delta lnk + v_L(\Delta lnL - \Delta lnH) + \Delta lnA, \end{aligned}$$

where lowercase letters are per hour worked.

Growth in ALP, Δlny, depends on three factors. The first is capital deepening, Δlnk, which captures the increase in capital services per hour. The second is the growth in labor quality, which measures substitution toward workers with higher marginal products and is defined as the difference between the growth of labor input and the growth of hours worked $(\Delta lnL - \Delta lnH)$. The third is the growth in TFP, ΔlnA, defined in equation 3, which captures the impact of technical change and other factors that raise output growth beyond the measured contribution of inputs.

If the neoclassical assumptions fail to hold, however, the Solow residual will not measure only technical change. Other factors that affect the Solow residual include distortions from imperfect competition, externalities and production spillovers, omitted inputs, cyclical fluctuations, nonconstant returns to scale, and reallocation effects. If there are increasing returns to scale but no technical change, for example, input shares will not equal output elasticities and one would estimate a positive Solow residual even though there is no technical change. While this may confound the interpretation of the Solow residual as a pure technology measure, it remains a useful indicator of the underlying technological forces. Basu and Fernald (1997), for example, report a high correlation between a traditional Solow residual and a more sophisticated index of technology that controls for market imperfections. Moreover, they argue that the Solow residual is an important welfare measure, even when it is not a measure of pure technical change.

Applications of the Neoclassical Model

Jorgenson and Stiroh (2000) use equations 2-4 in a traditional growth accounting analysis to study the sources of U.S. economic growth. They conclude that investment in tangible assets has been the dominant source of growth over the past four decades, while the contribution of technical progress as measured by the Solow residual has been relatively modest.

From 1959 to 1998, output grew 3.63 percent per year, reflecting 1.59 percent annual growth in hours and 2.04 percent growth in labor productivity (Table 1). Labor productivity growth reflects the contributions of capital deepening (1.10 percentage points per year), labor quality (0.32 percentage point per year), and TFP growth (0.63 percentage point per year). Thus, accumulation of tangible assets and human capital (measured as labor quality gains) has been an important part of U.S. output and productivity growth over the past four decades.

For the late 1990s, Jorgenson and Stiroh (2000) find that both an acceleration of TFP and rapid capital accumulation contributed to the recent U.S. productivity revival. For the 1995-98 period, estimated TFP growth of 0.99 percent per year was nearly three times as fast as it was during the 1973-95 period. Although the neoclassical model cannot really explain why TFP accelerated, it is nonetheless an important result since faster technical progress drives long-run productivity growth. Indeed, faster TFP growth is now incorporated into the rapid medium-run growth projections made by the Congressional Budget Office (2000).

Gauging the relative importance of capital deepening and technology has also been an important part of the debate surrounding the performance of the Asian newly industrialized countries (NICs). Krugman (1994), Young (1995), and Collins and Bosworth (1996) use this type of traditional neoclassical analysis to evaluate the potential for long-run growth in the NICs. All three conclude that broadly defined capital accumulation, as opposed to exogenous technical progress (measured as TFP growth), was the primary engine of growth for the NICs, and thus they are pessimistic about future growth prospects. Again, it is the neoclassical implication—that only technical change drives long-run productivity growth—that makes this distinction so important. These findings have led to a sharp debate over the relative importance of capital accumulation and TFP growth as sources of growth in these economies.[7]

These two applications of neoclassical growth accounting help to explain the proximate factors driving growth in the United States and in the NICs, although the conclusions must be kept in the proper perspective. As stressed by Hulten (1979), this methodology yields a valid measure of the rate of technical change if neoclassical assumptions hold, but also tends to understate the economic importance of it. For example, in a simple neoclassical model, faster technical change induces higher output, saving, investment, and capital accumulation, so part of historical capital accumulation itself is due to technical change in a deeper sense. It must be stressed, however, that the goal of growth accounting is to quantify the contribution of accumulated inputs correctly so that the rate of technical progress can be accurately measured. Modern growth accounting is more about measuring technical change than explaining it.

Table 1

Sources of U.S. Output and Productivity Growth, 1959-98

	1959-98	1959-73	1973-90	1990-95	1995-98
Growth in output (*Y*)	3.63	4.33	3.13	2.74	4.73
Growth in hours (*H*)	1.59	1.38	1.69	1.37	2.36
Growth in average labor productivity (*Y/H*)	2.04	2.95	1.44	1.37	2.37
Contribution of capital deepening	1.10	1.49	0.91	0.64	1.13
Contribution of labor quality	0.32	0.45	0.20	0.37	0.25
Contribution of aggregate total factor productivity	0.63	1.01	0.33	0.36	0.99

Source: Jorgenson and Stiroh (2000).

Notes: Decomposition of labor productivity growth is based on equation 4 in the text. All values are average annual percentages.

Moving beyond the Neoclassical Model

The appealing simplicity and intuition of the neoclassical framework have made it the backbone of applied work on productivity and economic growth.[8] Despite this popularity, however, several shortcomings make the standard neoclassical model not entirely satisfactory. First, early studies attributed the vast majority of labor productivity growth to exogenous forces.[9] Then, in the 1970s and 1980s, the neoclassical model failed to offer a persuasive explanation for important U.S. productivity trends like the post-1973 productivity slowdown. Second, since capital accumulation is subject to diminishing returns, steady-state growth in per capita variables inevitably grinds to a halt without exogenous technical progress. Finally, the international data did not seem to fit with the basic neoclassical model in terms of observed differences in income, capital shares and rates of return, and convergence properties.[10] These shortcomings led to several lines of subsequent research on the relationship between investment and productivity growth.

One approach, originated by Jorgenson and Griliches (1967) and summarized in Jorgenson (1996), remained firmly embedded in the neoclassical tradition and sought to develop better measures of investment, capital, labor, and other omitted inputs in order to reduce the magnitude of the unexplained residual. This approach did not seek to explain the origins of technical progress but rather to reduce its importance as an empirical explanation of growth. I return to the details of this work later. The second direction was the endogenous growth literature, to which I now turn.

The World of Endogenous Growth

The endogenous or new growth theory was developed to move beyond the neoclassical model by providing an endogenous mechanism for long-run productivity growth, either by removing the diminishing returns to capital or by explaining technical change as the result of specific actions. This literature is quite varied, and alternative explanations focus on many factors like different production structures, the dynamics of competition, innovation, increasing returns, and production spillovers. The following discussion describes a representative endogenous growth model.[11]

A Simple Endogenous Growth Model

A primary motivation for developing endogenous growth models was the desire to avoid the neoclassical implication that only exogenous technical progress drives long-run productivity growth. Indeed, one can simply assume a constant marginal product of capital as in the so-called "AK" models in which output is a linear function of capital, $y_t = Ak_t$.[12] In this case, long-run productivity growth can continue, and any change in the level of technology or savings rate leads to a long-run change in productivity growth.

Romer (1986), in a classic paper that sparked the new growth theory, provided a mechanism and corresponding economic explanation for why capital might not suffer from diminishing returns. In particular, Romer focused on the possibility of external effects as research and development efforts by one firm spill over and affect the stock of knowledge available to all firms. Firms face constant returns to scale to all private inputs, but the level of technology, A, varies, depending on the aggregate stock of some privately provided input

A primary motivation for developing endogenous growth models was the desire to avoid the neoclassical implication that only exogenous technical progress drives long-run productivity growth.

(5) $$Y_i = A(R) \cdot f(K_i, L_i, R_i),$$

where an i subscript represents firm-specific variables, R is the aggregate stock of knowledge, and time subscripts are dropped.[13]

The exact nature of the spillover is not particularly important for the aggregate properties of the model, and economists have identified several alternative channels. For instance, Arrow (1962) emphasizes "learning-by-doing," in which investment in tangible assets generates spillovers as aggregate capital increases; past gross investment proxies for experience and determines *A(.)*.[14] Alternatively, Romer (1986) essentially models *A(.)* as a function of the stock of R&D, Lucas (1988) models *A(.)* as dependent on the stock of human capital, and Coe and Helpman (1995) argue that *A(.)* also depends on the R&D stock of international trading partners. The key point is that there may be constant (or increasing) returns to accumulated inputs at the aggregate level, and thus the generation of long-run endogenous growth.[15]

The existence of production spillovers raises a significant empirical question that has generated a vast literature. If investment of any type—tangible assets, human capital, or R&D—generates benefits to the economy that are not internalized by private agents, then this suggests that there are multiple long-run growth paths and that there are specific policy implications. Since investment may be too low from society's point of view, spillovers open a role for government intervention such as investment tax credits or research and development grants. R&D spillovers, in particular, have attracted considerable attention in the new growth literature, and I review the microeconomic evidence on them later.

Macro Evidence of Endogenous Growth

One aggregate implication of early endogenous growth models is a "scale effect," in which productivity growth increases with the size of the economy. Larger economies devote more resources to R&D and knowledge production is available to all, so technology should grow more rapidly. In addition, this suggests that government policy, in the form of taxes or subsidies that increase resources allocated to knowledge production or investment, can raise long-run growth.

In a pair of influential studies, however, Jones (1995a, 1995b) strongly rejects this scale effect and finds little relationship between policy variables and long-run growth. There is no obvious relationship between the number of scientists and engineers employed in R&D activities and U.S. TFP growth (Jones 1995a), nor has there been persistent acceleration in growth in countries belonging to the Organization for Economic Cooperation and Development, even as investment shares rose dramatically (Jones 1995b). Over a very long horizon, however, the evidence of a scale effect is somewhat stronger; Kremer (1993) argues that productivity and population growth are highly correlated with initial population levels, as scale effects imply.

This influential critique has led to several alternatives to remove the link between scale and growth found in endogenous growth models.[16] Jones (1995a), for example, presents a model of "semi-endogenous" growth, in which long-run growth still reflects firms' R&D choices, but is independent of government policies such as investment tax credits or R&D subsidies. Scale or policy variables affect the levels of output and productivity but not long-run growth rates. The key factor determining long-run growth in this semi-endogenous growth model remains the degree of external returns in the R&D process, which delivers endogenous growth and provides the critical distinction from the neoclassical model.

In more recent work at the macro level, Jones and Williams (1998) formalize the macroeconomic impact of increasing overall returns due to various external effects related to R&D. They calibrate their model and estimate that optimal investment in R&D is two to four times actual investment in the United States. This work suggests an important role for R&D but remains consistent with the empirical refutation of the macro R&D models in Jones (1995a, 1995b).

Expanding the Investment Definition

I now turn to several broad measurement issues that directly affect how these types of models are implemented and evaluated. Over the years, sophisticated tools have been developed to measure inputs properly, and the definition of investment has been expanded beyond tangible assets. By quantifying substitution between heterogeneous tangible assets and explicitly recognizing investment in human capital, research and development, and public infrastructure, economists have made considerable methodological advances in the understanding of productivity growth. These improvements are now well-accepted and are part of the toolkits of most applied productivity analysts; for example, the official U.S. TFP estimates of the Bureau of Labor Statistics (2000a) account for asset substitution, human capital, and R&D.

These advances are also relevant to the debate between the neoclassical and new growth views. From the neoclassical view, improved measurement of inputs allows technical progress to be more accurately assessed. From the new growth view, constant returns may be more realistic for broader definitions of capital. The degree of economy-wide returns to accumulated inputs—diminishing returns in the neoclassical world and constant (or increasing) returns in the new growth world—remains the fundamental difference between the two views of economic growth, and this is conceptually independent of how broadly capital is defined or measured. Simply including additional inputs like human capital or R&D capital is not enough to generate endogenous growth if broadly defined capital still faces diminishing returns.

Finally, many of these methodological advances have been implemented within neoclassical growth accounting analyses. As mentioned above, the primary goal of these studies is to develop better measures of inputs and a more accurate estimate of technical progress. It is not, as is often assumed, to show that there is no technical progress. Careful growth accounting analyses purge the transitory impact of investment and input accumulation to leave behind a more precise estimate of the growth rate of technical progress.

Heterogeneous Tangible Assets

An important insight, first implemented by Jorgenson and Griliches (1967), is that one must account for the vast heterogeneity of capital inputs. A dollar spent on new computer equipment, for example, provides more productive services per period than a dollar spent on a new building. By explicitly recognizing these types of differences, one can estimate a "capital service flow" that is appropriate for the production function in equation 1.

Aggregate capital service flows are estimated by using asset-specific "user costs" or "rental prices" to weight each heterogeneous asset and to account for substitution between them. As firms respond to changing relative prices—for example, by substituting toward high-tech equipment and away from structures—a larger portion of investment is in assets with relatively high marginal products, and the capital service flow rises.[17] The difference between capital service flows and capital stock is called "capital quality" by some authors and reflects the changing composition of investment toward assets with higher marginal products.[18]

Jorgenson and Stiroh (2000) provide estimates of this decomposition between capital stock and capital quality. As shown in Table 2, capital stock accumulation has been the dominant force behind the growth in capital services in the United States, accounting for 1.32 of the 1.77 percentage point growth contribution of capital for the 1959-98 period. For the most recent period, the contribution of capital quality has increased markedly as firms steadily responded to changing relative prices and substituted toward high-tech equipment. Gordon (1999a) examines a longer time period, dating back to 1870, and concludes that quality adjustments of capital and labor inputs have been important sources of long-run growth in the United States.

The evidence shows that one must use a capital services methodology to quantify the role of capital accumulation correctly and to isolate technical progress. Failure to account for ongoing substitution understates the contribution of input accumulation and overstates the TFP residual from equation 3. Thus, this type of careful measurement of capital services is critical to obtaining accurate measures of technical progress and understanding long-run growth.

Human Capital

A second important extension of the capital concept is the explicit inclusion of the contribution of changes in labor quality. Economists have long recognized the importance of investments in human beings; and expenditures on education, job training, labor migration, and health care all increase the quality of human labor, enhance productivity, and are rightly called investments.[19]

Griliches (1960), Denison (1962), and Jorgenson and Griliches (1967) pioneered the use of wage data to weight heterogeneous workers and construct constant-quality indexes

Table 2
Growth Contribution of Capital and Labor Inputs, 1959-98

	1959-98	1959-73	1973-90	1990-95	1995-98
Contribution of capital services (*K*)	1.77	2.07	1.62	1.20	2.17
Contribution of capital stock	1.32	1.66	1.22	0.77	1.23
Contribution of capital quality	0.45	0.40	0.41	0.43	0.95
Contribution of labor input (*L*)	1.23	1.25	1.17	1.18	1.57
Contribution of labor hours	0.92	0.80	0.97	0.81	1.32
Contribution of labor quality	0.32	0.45	0.20	0.37	0.25

Source: Jorgenson and Stiroh (2000).

Notes: Growth contribution of capital quality equals the difference between the contribution of capital services and capital stock. Growth contribution of labor quality equals the difference between the contribution of labor input and labor hours. See the text for details. All values are average annual percentages.

of labor input. Similar to the treatment of capital, this approach captures substitution between different types of labor and results in a flow of labor inputs appropriate for the production function analysis of equation 1. In contrast, a simple sum of hours worked by heterogeneous workers ignores this type of compositional change.[20]

Using the framework in equations 3 and 4 and again defining labor quality as the difference in the growth of labor input and unweighted hours, Jorgenson and Stiroh (2000) estimate that labor quality growth accounted for nearly 15 percent of labor productivity growth for the 1959-98 period (Tables 1 and 2). Similarly, the Bureau of Labor Statistics (1999b) attributes one-third of U.S. nonfarm labor productivity growth from 1990 to 1997 to labor composition effects. Again, failure to account for these quality changes overstates historical TFP growth.

Looking at the impact of human capital from a different perspective, Mankiw, Romer, and Weil (1992) include investment in human capital in an augmented Solow growth model to examine cross-country differences in growth. Employing a Cobb-Douglas specification for aggregate output, they explicitly model human capital as a determinant of output as

$$Y = K^{\alpha}H^{\beta}(AL)^{1-\alpha-\beta}, \quad (6)$$

where H is the stock of human capital.[21]

Mankiw, Romer, and Weil (1992) use educational attainment to proxy for human capital accumulation. They find that this extended neoclassical model fits the data well in terms of the growth convergence predictions and the estimated output elasticities. The authors conclude that the augmented Solow model in its neoclassical form is consistent with the international evidence. In contrast, Lucas (1988) incorporates human capital in a growth model but explicitly includes an external spillover effect in the spirit of new growth theory.

Although these studies differ in their approach and questions—Jorgenson and Stiroh (2000) show labor quality to be an important contributor to U.S. productivity growth, and Mankiw, Romer, and Weil (1992) find human capital to be a good predictor of cross-country income differences—both emphasize the importance of accounting for human capital accumulation. By correctly identifying and measuring accumulated inputs, this extension of the neoclassical model allows for a better understanding of the growth process.

Research and Development

Knowledge creation through explicit research and development activities is a third extension of the capital accumulation process that deserves special attention. R&D, broadly defined as investment in new knowledge that improves the production process, has been the focus of considerable research activity. There is still some debate, however, over whether R&D is best viewed as a neoclassical factor of production or if it is best viewed as a source of spillovers as in the endogenous growth models.

It is straightforward to think of R&D as just another form of capital in which firms choose to invest. In this sense, R&D is not fundamentally different from investment in tangible capital. Griliches (1973, 1979), for example, argues that it is reasonable to view the primary impact of R&D investment in a neoclassical sense since returns accrue internally. Firms presumably invest in R&D to improve their own production processes and to raise profits, so any spillover effects are secondary and unintended consequences.

While it is conceptually straightforward to treat R&D as a neoclassical factor of production with diminishing returns, Griliches (1995), Hall (1996), and Jorgenson (1996) all emphasize the practical difficulty in measuring the growth contribution of R&D because of thorny measurement problems and a lack of adequate data. Hall (1996) points out that R&D is often associated with product improvements, and the measured impact of R&D therefore depends critically on how price deflators are constructed and how output is deflated. In addition, one must estimate an appropriate depreciation rate to calculate the productive stock of R&D capital.

Despite these problems, many studies have attempted to measure the direct impact of R&D.[22] As an example of the typical approach, Griliches (1995) presents a skeletal model of R&D that is a straightforward extension of the neoclassical framework:

$$lnY = a(t) + \beta lnX + \gamma lnR + \varepsilon, \quad (7)$$

where X is a vector of standard inputs, such as capital and labor, and R is a measure of cumulative research effort.

Studies of this type typically have found that R&D capital contributes significantly to cross-sectional variation in productivity. It is important to emphasize that equation 7 examines the relationship between firm or industry productivity and its own R&D stock. In the equation, R&D is treated as a conventional neoclassical capital input with internal rewards.

Others, however, argue that R&D capital is fundamentally different from tangible and human capital. Knowledge capital appears to be noncompetitive since many producers can use the same idea simultaneously, and the returns may be hard to appropriate due to potential production spillovers. This difference is what makes R&D capital an important part of the endogenous growth models discussed earlier.[23] As emphasized by Romer (1994) and Basu (1996), the distinction between internal and external benefits drives the difference in returns to capital that delineates the role of research and development in the neoclassical and the new growth theories.

There seems to be some confusion, however, about what constitutes a true production spillover and what is really a more conventional measurement problem. Griliches (1995, p. 66) defines a production (knowledge) spillover as "ideas borrowed by research teams of industry *i* from the research results of industry *j*." This is quite distinct from situations in which transaction prices do not fully reflect the marginal benefit of the innovation (for example, see Bresnahan [1986], Bartelsman, Caballero, and Lyons [1994], and Keller [1998]). Similarly, Hall (1996) discusses how competition may lead to lower prices for goods of innovative firms. Rather than measuring spillovers in Griliches' sense, these gains to innovation reflect the inaccuracy of prices that do not adequately capture changes in the quality dimension.[24] While there are daunting practical difficulties, if all attributes and quality characteristics could be correctly priced, then increased quality or variety of intermediate inputs would not be measured as productivity spillovers.

The distinction between internal and external benefits drives the difference in returns to capital that delineates the role of research and development in the neoclassical and the new growth theories.

If true productivity gains do spill over to other firms, however, one channel for endogenous growth is opened. The microeconomic evidence suggests that R&D spillovers may exist,[25] but a wide variation in results and conceptual difficulties suggest that some caution is warranted. Griliches (1995), for example, points out that the impact of R&D in industry analyses is not greater than in firm analyses (as the presence of spillovers implies) and warns, "in spite of a number of serious and promising attempts to do so, it has proven very difficult to estimate the indirect contribution of R&D via spillovers to other firms, industries and countries" (p. 83). Given the poor quality of the data and methodological problems discussed earlier, it is difficult to draw definitive conclusions from these studies.

Public Infrastructure

A final extension of the investment concept worth noting is public infrastructure investment. In a series of influential and controversial studies, Aschauer (1989a, 1989b, 1990) argues that core infrastructure investment is an important source of productivity growth and that the sluggish productivity performance of the 1970s can be largely attributed to a slowdown in public investment. These claims led to a wide-ranging debate about policy implications and possible methodological problems such as potential biases from common trends, omitted variables, and potential reverse causality.[26]

Independent of any methodological concerns, the primary impact of government capital is conceptually the same as that of tangible capital and depends on proper measurement.[27] In a standard neoclassical growth accounting framework with only private inputs, any impact from government capital would be mismeasured as TFP growth. In this sense, government capital is just another accumulated input that is often missed and thus contributes to an overstatement of true technical change. As long as diminishing returns to all capital exist, the neoclassical implications hold.

Alternatively, Barro (1990) suggests that government services generate constant returns to broadly defined capital and lead to endogenous growth. In this view, government capital differs from private capital, and the real question is whether long-run constant returns to scale exist across broadly defined capital.[28]

The recent results of Fernald (1999) shed some light on this question. He shows that investment in roads contributed substantially to productivity prior to 1973, but he also suggests that new investment in roads offers a normal or even zero return at the margin. That is, the original interstate highway system improved productivity, but a second one would not. While not an exact test, this seems more consistent with the neoclassical view of diminishing returns of capital than with the endogenous growth view.

Recent Productivity Controversies

I now move to a discussion of two current and important issues relating to investment, productivity, and growth. Both are concerned with understanding the sources of productivity growth and draw on the neoclassical and endogenous growth theories described above. In particular, I offer a resolution to the computer productivity paradox and review the renewed embodiment controversy.

The Computer Productivity Paradox

Over the past few decades, investment in high-tech equipment, particularly computers, exploded, but aggregate productivity growth remained sluggish through the mid-1990s. This apparent contradiction—the so-called computer productivity paradox of the 1980s and early 1990s—disappointed many observers and initiated a broad research effort at both the macro and micro levels. More recently, however, productivity growth has accelerated sharply (see chart); I argue that this pattern is entirely consistent with neoclassical explanations of capital accumulation and technical progress.[29]

The defining characteristic of the information technology revolution is the enormous improvement in the quality of computers, peripherals, software, and communications equipment. As epitomized by Moore's Law—the doubling of the power of a computer chip every eighteen months—each generation of new computers easily outperforms models considered state-of-the-art just a few years earlier. The constant-quality deflators employed by the U.S. Bureau of Economic Analysis translate these massive quality improvements into increased real investment and real output and show annual price declines of 18 percent over the past four decades.[30]

How does this IT phenomenon fit within the neoclassical framework? To address this question, it is critical to distinguish between the *use* and the *production* of IT.[31] Information technology is both an output from the IT-producing industries and an input to the IT-using industries, so there are two effects. The massive quality improvements in IT contribute to faster productivity growth in the IT-producing industries and faster input accumulation in the IT-using industries. Thus, the neoclassical model predicts rapid capital deepening and ALP growth in IT-using industries, and technical progress and TFP growth in the IT-producing industries. This fundamental distinction is apparent in Solow (1957), but often has been overlooked in discussions of the computer productivity paradox.

IT Use, Capital Deepening, and Productivity Growth

Consider the productivity of firms and industries that invest in and use information technology. Following the neoclassical framework in equation 4, strong IT investment contributes directly to ALP growth through traditional capital deepening effects. In the case of IT, this reflects rapid growth in capital services and, until recently, a small income share. Over the past four decades, however, U.S. firms have continued to respond to large price declines by investing heavily, particularly in computer hardware, and rapidly accumulating IT.

As long as relative prices continued to fall, it was inevitable that IT inputs would eventually make a large contribution to growth. Indeed, recent estimates by Jorgenson and Stiroh (2000), Oliner and Sichel (2000), and Whelan (2000) indicate substantial increases in the growth contributions from computers and other IT capital. Jorgenson and Stiroh (2000), as reported in Table 3, estimate that the contribution of computer hardware increased from 0.19 percentage point per year for the 1990-95 period to 0.46 percentage point for 1995-98; Oliner and Sichel (2000) report an increase from 0.25 for 1990-95 to 0.63 for 1996-99.[32] Both agree that IT capital accumulation has been an important part of the acceleration in U.S. productivity since 1995.

Table 3
Growth Contribution of Information Technology and Other Assets, 1959-98

	1959-98	1959-73	1973-90	1990-95	1995-98
Contribution of capital services (*K*)	1.77	2.07	1.62	1.20	2.17
Other capital	1.45	1.89	1.27	0.80	1.42
Computer hardware	0.18	0.09	0.20	0.19	0.46
Computer software	0.08	0.03	0.07	0.15	0.19
Communications equipment	0.07	0.06	0.08	0.06	0.10

Source: Jorgenson and Stiroh (2000).

Notes: A growth contribution is defined as the share-weighted real growth rate of the asset as in equation 3 in the text. All values are average annual percentages.

This historical record on IT capital accumulation appears entirely consistent with the neoclassical model. Massive input substitution and rapid capital accumulation during the 1990s have led to an aggregate growth contribution that is now quite large. Prior to that, the contribution was modest because the stock of IT was small. Only in the late 1990s, after a major information technology investment boom, were there enough information technology inputs to have a substantial impact on growth and labor productivity at the aggregate level.

Despite the straightforward relationship and mounting aggregate evidence, the empirical evidence on computers and ALP growth across industries has been mixed. Gera, Gu, and Lee (1999), McGuckin, Steitwieser, and Doms (1998), McGuckin and Stiroh (1998), and Steindel (1992) find a positive impact, while Berndt and Morrison (1995) report a negative impact. Morrison (1997) finds overinvestment in high-tech assets for much of the 1980s. The microeconomic evidence for firms is somewhat stronger. For example, Brynjolfsson and Hitt (1993, 1995, 1996), Lehr and Lichtenberg (1999), and Lichtenberg (1995) typically estimate returns to computers that exceed other forms of capital.

Only in the late 1990s, after a major information technology investment boom, were there enough information technology inputs to have a substantial impact on growth and labor productivity at the aggregate level.

On the surface, exceptionally high returns to IT suggest effects found in some endogenous growth models. The findings of large gross returns, however, are also quite consistent with the neoclassical model—computers must have high marginal products because they obsolesce and lose value so rapidly. Computers have a low acquisition price, but rapid obsolescence makes them expensive to use and a high return is needed as compensation. This is exactly the type of asset heterogeneity for which the user-cost methodology was designed. In addition, Brynjolfsson and Yang (1997) suggest that much of the "excess returns" to computers actually represent returns to previously unspecified inputs such as software investment, training, and organizational change that accompany computer investment.

Note that the neoclassical framework predicts no TFP growth from IT use since all output contributions are due to capital accumulation. Computers increase measured TFP only if there are nontraditional effects like increasing returns, production spillovers, or network externalities, or if inputs are measured incorrectly. The nontraditional effects, if present, would move the IT revolution into the world of new growth theory. The evidence, however, is not very strong. Siegel and Griliches (1992) and Siegel (1997) estimate a positive impact of computer investment on TFP growth across U.S. industries, while Berndt and Morrison (1995) and Stiroh (1998a) do not. Working with aggregate data, Gordon (2000) finds no evidence of computer-related spillovers in the late 1990s. In sum, there does not appear to be compelling evidence for nontraditional effects from IT that lead outside of the neoclassical model.

IT Production and Technical Progress

Now consider the productivity of firms and industries that produce IT assets. These sectors are experiencing fundamental technical progress—the ability to produce more output in the form of faster processors, expanded storage capacity, and increased capabilities from the same inputs—that should be measured directly as industry TFP and lead to higher ALP growth (equations 3 and 4). This affects both industry and aggregate productivity.

The data are quite clear that fundamental technical progress, measured as TFP growth, is a driving force in the production of these new high-tech assets. The Bureau of Labor Statistics (1999b), Jorgenson and Stiroh (2000), Oliner and Sichel (2000), and Stiroh (1998a) all report strong industry TFP growth in the high-tech producing industries. Moreover, much of the acceleration in aggregate U.S. productivity growth after 1995 can be traced to accelerations in the pace of technical progress—measured as faster relative price declines in these high-tech industries.[33]

This notion that technical progress in specific industries drives aggregate productivity is hardly new and is consistent with the broad neoclassical framework. As early as Domar (1961), economists recognized that aggregate TFP growth reflects technical progress among component industries. Accelerating technical progress in key industries can then drive aggregate productivity through both a direct TFP contribution and induced capital accumulation as relative prices change.

Alternative Explanations

The preceding discussion still leaves open the question of why even ALP growth remains sluggish in some computer-intensive sectors like finance, insurance, real estate, and services. Since computers are highly concentrated in service sectors, where output and productivity are notoriously hard to measure, Diewert and Fox (1999), Griliches (1994), Maclean (1997), and McGuckin and Stiroh (1998) suggest that measurement error plays a role in the remaining computer productivity paradox for certain IT-using industries.

A second common explanation is that computers are still relatively new and it may just be a matter of time until they fundamentally change the production process and usher in a period of faster productivity growth throughout the economy. David (1989, 1990) draws a parallel between the slow productivity benefits from electricity and those from computers. Triplett (1999a), however, cautions against such analogies, arguing convincingly that the massive declines in computer prices, and hence the diffusion patterns, are unprecedented. Moreover, computers are no longer really a new investment—the first commercial purchase of a UNIVAC mainframe computer was in 1954, and computer investment has been a separate entity in the U.S. national accounts since 1958.[34] This critical mass and delay hypothesis is beginning to lose credibility as an explanation for low productivity in certain computer-intensive industries.

Technical change in the production of information technology assets lowers the relative price, induces massive high-tech investment, and is ultimately responsible for the recent productivity revival.

A final explanation is simply that computers are not that productive in some industries. Anecdotes abound of failed systems, lengthy periods of downtime, unwanted and unnecessary "features," and time-consuming upgrades—all of which can reduce the productivity of computer investment.[35]

IT and the New Economy

Despite some lingering questions, the computer productivity paradox appears to be over. Aggregate productivity growth is strong, and IT-producing industries are showing rapid TFP growth. Moreover, the IT revolution and the "new economy" appear to be largely a neoclassical story of relative price declines and input substitution. Technical change in the production of information technology assets lowers the relative price, induces massive high-tech investment, and is ultimately responsible for the recent productivity revival.

These benefits, however, accrue primarily to the producers and users of IT, with little evidence of large spillovers from computers. That is, we see TFP growth in IT-producing industries and capital deepening elsewhere. Of course, the neoclassical model provides no explanation for why technical progress may have accelerated in high-tech industries in recent years. Perhaps models of endogenous innovation and competition can provide an answer.

The Renewed Embodiment Controversy

The discussion so far has focused on the modern neoclassical framework and new growth models as explanations of productivity growth. An alternative perspective, however, argues that technological progress is "embodied" in new machinery and equipment, as opposed to being a more pervasive force that affects the production of all goods and services. In challenging papers, Greenwood, Hercowitz, and Krusell (1997) and Hercowitz (1998) recently brought this debate back to center stage and reopened the embodiment controversy.

This debate is important since it helps us to understand the precise roles of technology and investment in the growth process. By better understanding these issues, such as whether technology is driving the recent investment boom, one may be able to implement more effective policies. Moreover, there are clear implications associated with how real aggregate output should be measured that directly affect how the economy is viewed. For example, Greenwood et al. (1997) disagree with the current practice of adjusting the output of investment goods, such as computers, for quality improvements. As noted in the previous section, gains in the production of computers have been a major part of the productivity revival of the 1990s and excluding them would substantially change our perception of the U.S. economy.

The embodiment idea goes back at least to Solow (1960), who suggests that technical change is embodied in new investment goods, which are needed to realize the benefits of technical progress. In response, Jorgenson (1966) shows this to be no different conceptually from the neoclassical view of disembodied technical change with the calculation of investment price deflators responsible for apparent differences. If new vintages of capital have different productive characteristics, the appropriate constant-quality deflators attribute output and productivity growth to input accumulation and not to technical progress. Hulten (1992) shows how failure to account for improved characteristics of recent vintages of investment goods suppresses capital enhancements into the traditional TFP residual.

There seems to be general agreement that different vintages of capital inputs need to be adjusted for quality change in order to understand and quantify the sources of growth. A second conclusion of Jorgenson (1966), however, is that investment as an output must also be measured in quality-adjusted units; Hulten (2000) discusses potential measurement errors when investment is not measured in such units. While this methodology has been integrated into the U.S. national accounts—where constant-quality deflators are used for investment goods like computer hardware and software to calculate real GDP—it plays a central role in the renewed embodiment debate.

An alternative perspective [to the neoclassical and new growth models] argues that technological progress is "embodied" in new machinery and equipment.

Greenwood et al. (1997) and Hercowitz (1998) argue explicitly against adjusting investment output for quality change, preferring to measure real investment output in units of consumption. As motivating evidence, Greenwood et al. report that the relative price of equipment in the United States has fallen 3 percent per year in the postwar era, while the equipment-to-GDP ratio increased dramatically. They calibrate a balanced growth path and attribute 60 percent of postwar productivity growth to "investment-specific technological change" that is conceptually distinct from capital accumulation and disembodied (Hicks-neutral) technological change. A clear implication, they argue, is that constant-quality price indexes are appropriate only for deflating investment inputs, not for deflating investment as an output.[36]

It is essential, in my view, that both investment outputs and capital inputs be measured consistently in quality-adjusted units, as currently employed in the U.S. national accounts. The key point is that improved production characteristics of new investment cohorts are themselves produced and rightly considered output. Although a rigorous theoretical model is beyond the scope of this article, it is useful to sketch out a defense of this position.

Consider a simple two-sector, neoclassical model like the one outlined in the earlier discussion of the computer productivity paradox. One sector produces computers and one sector uses computers; both use neoclassical production functions. Can this explain the motivating evidence cited by Greenwood et al.? The answer is clearly yes. Disembodied technical change in the computer-producing industry (measured as industry-level TFP) reduces the marginal cost of producing computers and lowers prices. Profit-maximizing firms elsewhere in the economy respond to the relative price change, substitute between inputs, and rapidly accumulate computers. Thus, one would observe falling relative prices and rising investment shares in a traditional neoclassical world with purely disembodied technical change in one sector.

Greenwood et al. appear to need investment-specific technical change since they focus on a one-sector model in which consumption, equipment investment, and construction equipment are produced from the same aggregate production function. This effectively imposes perfect substitutability between investment and consumption goods and requires investment-specific technical change to explain relative price changes. In a multisector neoclassical model like the one proposed by Domar (1961) and implemented by Jorgenson et al. (1987), investment-specific technical change is not needed since disembodied technical progress can proceed at different rates in different industries and generate the observed relative price changes.[37]

What does this say about the appropriate deflation of investment goods? In this example, the answer clearly depends on what one believes the computer-producing sector actually produces. High-tech industries expend considerable resources in the form of investment, R&D, and labor to produce better, faster, and more powerful products. Indeed, the hedonic literature on computers is based on the idea that the computer-producing industry creates computing power measured as a bundle of productive characteristics, rather than as computer boxes or units.[38] If computing power measured in quality-adjusted efficiency units is the appropriate measure of industry output, and I believe it is, then internal accounting consistency requires that aggregate output also be measured in quality-adjusted units.

Finally, and perhaps more fundamentally, the neoclassical model at its most basic level is a model of production and not of welfare. This too implies that output must be a measure of produced characteristics in terms of quality-adjusted units, rather than consumption forgone. The embodiment approach, in contrast, confounds the link between the sources of growth (labor, capital, and technology) and the uses of growth (consumption and investment goods) that constitute separate views of production and welfare.

Conclusion

This article provides a broad overview of the link between investment and productivity in two alternative views—the neoclassical and new growth models. Although the models emphasize different aspects of productivity growth, they both contribute to our understanding of the growth process.

The key distinction between the neoclassical and new growth theories concerns the aggregate returns to capital and the implications for long-run productivity growth. In the neoclassical world, capital (broadly defined to include all accumulated inputs) suffers from diminishing returns, and productivity growth is ultimately determined by exogenous technical progress. In the world of endogenous growth, there can be constant returns to capital that generate long-run growth in per capita variables. Although both views attempt to explain growth, they focus on different aspects and need not be mutually exclusive. Neoclassical economists developed sophisticated measurement tools to identify technical progress accurately by removing the transitory impact of input accumulation; new growth theorists developed sophisticated growth models to explain the evolution of technology as a result of the actions of economic agents. Both contributions are important.

Attempts to understand recent changes in the U.S. economy illustrate this complementarity. Aggregate productivity growth has accelerated in the past few years due to a combination of accelerating technical progress in high-tech industries and corresponding investment and capital deepening. This type of neoclassical analysis clearly explains *what* happened to the U.S. economy. To explain *why* it happened, we need to focus on the incentives and actions of the firms that actually invest, innovate, and create the new capital and knowledge that are driving the U.S. economy. This is the domain of the endogenous growth framework. Thus, both approaches make important contributions to our understanding of the economic growth process.

Endnotes

1. Labor productivity is defined as real output per hour worked. These estimates are for the U.S. nonfarm business sector from the Bureau of Labor Statistics (2000b) and are consistent with the revised GDP data after the benchmark revision in October 1999.

2. See, for example, Jorgenson and Stiroh (2000), Oliner and Sichel (2000), Gordon (1999b, 2000), and Parry (2000).

3. See Jorgenson (1996) for a discussion of the growth theory revival, Barro and Sala-i-Martin (1995) for a thorough analysis of the neoclassical framework, and Aghion and Howitt (1998) for a detailed review of different strands of new growth theory. The terms "new growth" and "endogenous growth" are used interchangeably throughout this article.

4. The Solow model assumes labor is fully employed so per capita income and labor productivity growth coincide.

5. The neoclassical model has also been used extensively to examine cross-country differences in the growth and level of output. This vast body of literature is not discussed here; see Barro and Sala-i-Martin (1995) for references and Mankiw (1995) for a summary of the strengths and weaknesses of the neoclassical model in this context.

6. Under the neoclassical assumptions, an input's elasticity equals its share of nominal output since the marginal product of an input equals its factor price, for example, the wage rate for labor and the rental price for capital.

7. Hsieh (1997, 1999), Rodrick (1997), and Young (1998b) provide recent views on this controversy.

8. Stiroh (1998b) reports that the long-run projection models used by various U.S. government agencies—for example, the Social Security Administration, the Congressional Budget Office, the Office of Management and Budget, and the General Accounting Office—are all firmly embedded in this neoclassical tradition.

9. Solow (1957), for example, estimates that nearly 90 percent of per capita income growth is due to technical progress.

10. See Mankiw (1995), particularly pp. 280-9.

11. My working definition follows Segerstrom (1998), who defines endogenous growth models as those in which "the rates of technological change and economic growth are endogenously determined based on the optimizing behavior of firms and consumers" (p. 1292). Hulten (2000) identifies noncompetitive markets, increasing returns to scale, externalities, and endogenous innovation as the key aspects of the new growth theory.

12. Technically, long-run growth in per capita variables exists under constant returns to all accumulated inputs. Note that in the simplest AK model like this one, A represents a constant level of technology, in contrast to the general production functions in the text.

13. These simplifications follow Romer (1994), who summarizes the evolution of endogenous growth models. One alternative mechanism is to allow for increasing returns at the level of individual firms. However, this approach is inconsistent with perfect competition. See Aghion and Howitt (1998) for a discussion. In addition, there are aggregation concerns when moving from a firm to an aggregate production function.

14. Note that Arrow (1962) does not explicitly derive a model of endogenous growth. In his model, growth eventually stops if population is held constant.

15. Barro (1990) achieves endogenous growth in a model with constant returns to capital and government services together, but diminishing returns to private capital alone. DeLong and Summers (1991, 1992, 1993), although not modeling endogenous growth, argue that equipment investment yields large external benefits in the spirit of Arrow (1962).

16. See, for example, Jones (1995a), Kortum (1997), Segerstrom (1998), and Young (1998a). Jones (1999) reviews.

17. An asset's rental price reflects the opportunity cost of buying the asset, depreciation, and any capital gains or losses on the asset. High-tech equipment experiences more rapid depreciation and smaller capital gains than structures, so equipment must have a correspondingly higher marginal product and service price. See Hall and Jorgenson (1967) for the original derivation and Jorgenson and Stiroh (2000) for a recent application and details.

18. Note that capital quality in this framework does not reflect increased productive power from a particular asset. These gains, such as the enhanced performance of more recent vintages of computers, are handled by the investment deflator, which translates nominal investment into larger quantities of real investment. I provide details later in this article.

19. Mincer (1958, 1974), Shultz (1961), and Becker (1962) are early examples; Griliches (1996) provides a summary of the early work on human capital. As early as 1961, the similarities between investments in tangible capital and human capital such as tax incentives, depreciation, pricing imperfections, and the primarily internal benefits of human capital investments were discussed by Schultz (1961, pp. 13-5).

20. This methodology provides an index of aggregate human capital that changes as the composition of the labor force changes. The key assumption is that the quality of a particular type of labor—for example, a person of a given age, education, experience—is constant over time. Ho and Jorgenson (1999) provide methodological details.

21. Note that Mankiw, Romer, and Weil (1992) explicitly assume there are diminishing returns to accumulated inputs, $\alpha + \beta < 1$, which places the model squarely in the neoclassical tradition.

22. Good et al. (1996), Griliches (1994, 1995), and Hall (1996) provide detailed surveys of the empirical literature.

23. Hall (1996) offers a number of reasons why R&D might lead to production spillovers such as reverse engineering, migration of scientists and engineers, and free dissemination of public R&D. Grossman (1996, particularly pp. 86-8) emphasizes the differences between R&D capital and tangible capital.

24. See Griliches (1992) for a discussion of this distinction.

25. Good et al. (1996) state that “most of these recent studies point in the direction that there is some effect of R&D spillovers on the productivity growth of the receiving industry or economies” (p. 39). Griliches (1992) states that “in spite of many difficulties, there has been a significant number of reasonably well-done studies all pointing in the same direction: R&D spillovers are present, their magnitude may be quite large, and social rates of return remain significantly above private rates” (p. S43).

26. The conference proceedings in Munnell (1990) explore this issue. Aaron (1990) is a good example of important critiques of the Aschauer findings. Gramlich (1994) and Binder and Smith (1996) provide more recent reviews.

27. Measurement problems may be more severe for government capital because there are no markets for many types of such capital, which makes estimation of user costs difficult.

28. One obvious difference between private and public investment is the financing mechanism. For example, the typical argument for government infrastructure investment is a traditional public-good argument that prevents returns from being recouped by the investor, which can lead to underprovision of the good. Gramlich (1990) discusses this in detail.

29. Brynjolfsson and Yang (1996) summarize earlier empirical work, Sichel (1997) provides a broad analysis of the impact of computers, and Triplett (1999a) presents a detailed critique of common explanations for the productivity paradox. Ultimately, one would like to answer a difficult counterfactual question—how fast would labor productivity have grown in the absence of computers?—but this is very difficult indeed. For example, the explosion of computing power occurred roughly contemporaneously with the well-known productivity slowdown, and one must distinguish the productivity impact of computers from the host of factors examined in that context. See Federal Reserve Bank of Boston (1980), Baily and Gordon (1988), Baily and Schultze (1990), and Wolff (1996) for a few examples of the extensive literature on the productivity slowdown.

30. There is strong agreement that adjusting the output of computers for quality change is appropriate, but there are dissenting views. Denison (1989) argues specifically against constant-quality price indexes for computers.

31. See Baily and Gordon (1988), Stiroh (1998a), Gordon (2000), Jorgenson and Stiroh (2000), and Oliner and Sichel (2000) for details on this fundamental distinction. This discussion is also relevant to the discussion of the renewed embodiment controversy below.

32. These empirical differences primarily reflect the time periods and output concepts. See Oliner and Sichel (2000) for details.

33. As an important caveat, Triplett (1996) shows that one must incorporate quality adjustments for all inputs to correctly allocate TFP across sectors to specific high-tech industries. His results suggest that falling prices of computers are in large part due to enormous technical progress in the production of semiconductors, a crucial intermediate input to computer production. See Oliner and Sichel (2000) for recent estimates.

34. Gordon (1989) provides a history of the early evolution of computers.

35. Gordon (2000) summarizes this pessimistic view. Kiley (1999) presents a formal model of how computer investment could lower productivity due to large adjustment costs.

36. Reported relative price changes are based on Gordon (1990) and extended forward. This is a puzzling appeal to evidence, however, since the goal of Gordon's monumental effort was to develop better output price measures and he explicitly states, "both input price and output price indexes treat quality change consistently" (p. 52). Moreover, this approach assumes no quality change in consumption goods, so measured relative price changes are more accurately thought of as the relative rate of technical change between these two.

37. In their introduction, Greenwood et al. (1997) seem to agree; they view their motivating facts as evidence of "significant technological change in the production of new equipment" (p. 342). Although they do calibrate a simple two-sector model, it is not fully integrated with their empirical work on the sources of growth and they reject it as an unreasonable explanation of balanced growth rates.

38. See Triplett (1989) for a survey.

References

Aaron, Henry J. 1990. Discussion of "Why Is Infrastructure Important?" In Alicia H. Munnell, ed., Is There a Shortfall in Public Capital Investment? Federal Reserve Bank of Boston Conference Series no. 34: 51-63.

Aghion, Philippe, and Peter Howitt. 1998. Endogenous Growth Theory. Cambridge, Mass.: MIT Press.

Arrow, Kenneth J. 1962. "The Economic Implications of Learning by Doing." Review of Economic Studies 29 (June): 155-73.

Aschauer, David. 1989a. "Is Public Expenditure Productive?" Journal of Monetary Economics 23, no. 2: 177-200.

———. 1989b. "Does Public Capital Crowd out Private Capital?" Journal of Monetary Economics 24, no. 2: 171-88.

———. 1990. "Why Is Infrastructure Important?" In Alicia H. Munnell, ed., Is There a Shortfall in Public Capital Investment? Federal Reserve Bank of Boston Conference Series no. 34: 21-50.

Baily, Martin Neil, and Robert J. Gordon. 1988. "The Productivity Slowdown and Explosion of Computer Power." Brookings Papers on Economic Activity, no. 2: 1-45.

Baily, Martin Neil, and Charles L. Schultze. 1990. "The Productivity of Capital in a Period of Slower Growth." Brookings Papers on Economic Activity, Microeconomics (Special Issue): 369-406.

Barro, Robert J. 1990. "Government Spending in a Simple Model of Endogenous Growth." Journal of Political Economy 98, no. 5 (part 2): S103-25.

Barro, Robert J., and Xavier Sala-i-Martin. 1995. Economic Growth. New York: McGraw-Hill.

Bartelsman, Eric J., Ricardo J. Caballero, and Richard K. Lyons. 1994. "Customer- and Supplier-Driven Externalities." American Economic Review 84, no. 4 (September): 1075-84.

Basu, Susanto. 1996. Discussion of "Technology in Growth Theory." In Jeffrey C. Fuhrer and Jane Sneddon Little, eds., Technology and Growth. Federal Reserve Bank of Boston Conference Series no. 40: 78-82.

Basu, Susanto, and John G. Fernald. 1997. "Aggregate Productivity and Aggregate Technology." Board of Governors of the Federal Reserve System International Finance Discussion Paper no. 593.

Becker, Gary. 1962. "Investment in Human Capital: A Theoretical Analysis." Journal of Political Economy 70, no. 5 (part 2): 9-49.

Berndt, Ernst R., and Catherine J. Morrison. 1995. "High-Tech Capital Formation and Economic Performance in U.S. Manufacturing Industries: An Exploratory Analysis." Journal of Econometrics 65, no. 1: 9-43.

Binder, Susan J., and Theresa M. Smith. 1996. "The Linkage between Transportation Infrastructure Investment and Productivity: A U.S. Federal Research Perspective." In David F. Batten and Charlie Karlsson, eds., Transportation Infrastructure Investment and Productivity. Heidelberg, Germany: Springer.

Bresnahan, Timothy F. 1986. "Measuring the Spillovers from Technical Advance: Mainframe Computers in the Financial Services." American Economic Review 76, no. 4 (September): 741-55.

Brynjolfsson, Erik, and Lorin Hitt. 1993. "Is Information Systems Spending Productive? New Evidence and New Results." Proceedings of the International Conference on Information Systems, December.

———. 1995. "Information Technology as a Factor of Production: The Role of Differences among Firms." Economics of Innovation and New Technology 3, nos. 3-4: 183-99.

———. 1996. "Paradox Lost? Firm-Level Evidence on the Returns to Information Systems Spending." Management Science 42, no. 4 (April): 541-58.

Brynjolfsson, Erik, and Shinkyu Yang. 1996. "Information Technology and Productivity: A Review of the Literature." Advances in Computers 43 (February): 179-214.

———. 1997. "The Intangible Benefits and Costs of Computer Investments: Evidence from the Financial Markets." Proceedings of the International Conference on Information Systems, December.

References (Continued)

Bureau of Labor Statistics. 1998. “International Comparisons of Manufacturing Productivity and Unit Labor Cost Trends, 1997.” U.S. Department of Labor News Release no. 98-393, September 25.

———. 1999a. “Productivity and Costs, Preliminary Fourth-Quarter Measures and Annual Averages, 1998.” U.S. Department of Labor News Release no. 99-32, February 9.

———. 1999b. “Multifactor Productivity Trends.” U.S. Department of Labor News Release no. 99-36, February 11.

———. 2000a. “Multifactor Productivity Trends, 1998.” U.S. Department of Labor News Release no. 00-267, September 21.

———. 2000b. “Productivity and Costs, First Quarter 2000.” U.S. Department of Labor News Release no. 00-164, June 6.

Coe, David T., and Elhanan Helpman. 1995. “International R&D Spillovers.” European Economic Review 39, no. 5: 859-87.

Collins, Susan M., and Barry P. Bosworth. 1996. “Economic Growth in East Asia: Accumulation versus Assimilation.” Brookings Papers on Economic Activity, no. 2: 135-203.

Congressional Budget Office. 2000. “The Budget and Economic Outlook: Fiscal Years 2001-2010.” Washington, D.C.: U.S. Government Printing Office.

David, Paul A. 1989. “Computer and Dynamo: The Modern Productivity Paradox in a Not-Too-Distant Mirror.” Center for Economic Policy Research Publication no. 172.

———. 1990. “The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox.” American Economic Review 80, no. 2 (May): 35-61. Papers and Proceedings of the 102nd Annual Meeting of the American Economic Association.

DeLong, J. Bradford, and Lawrence H. Summers. 1991. “Equipment Investment and Economic Growth.” Quarterly Journal of Economics 106, no. 2: 445-502.

———. 1992. “Equipment Investment and Economic Growth: How Strong Is the Nexus?” Brookings Papers on Economic Activity, no. 2: 157-99.

———. 1993. “How Strongly Do Developing Countries Benefit from Equipment Investment?” Journal of Monetary Economics 32, no. 3: 395-415.

Denison, Edward F. 1962. Sources of Economic Growth in the U.S. and the Alternatives before Us. New York: Committee for Economic Development.

———. 1989. Estimates of Productivity Change by Industry. Washington, D.C.: Brookings Institution.

Diewert, Erwin, and Kevin Fox. 1999. “Can Measurement Error Explain the Productivity Paradox?” Canadian Journal of Economics 32, no. 2 (April): 251-80.

Domar, Evsey. 1961. “On the Measurement of Technical Change.” Economic Journal 71, no. 284 (December): 709-29.

Federal Reserve Bank of Boston. 1980. “The Decline in Productivity Growth.” Federal Reserve Bank of Boston Conferences Series no. 22.

Fernald, John G. 1999. “Roads to Prosperity? Assessing the Link between Public Capital and Productivity.” American Economic Review 89, no. 3: 619-38.

Gera, Surendra, Wulong Gu, and Frank C. Lee. 1999. “Information Technology and Labour Productivity Growth: An Empirical Analysis for Canada and the United States.” Canadian Journal of Economics 32, no. 2 (April): 384-407.

Good, David H., M. Ishaq Nadiri, and Robin C. Sickles. 1996. “Index Number and Factor Demand Approaches to the Estimation of Productivity.” NBER Working Paper no. 5790.

Gordon, Robert J. 1989. “The Postwar Evolution of Computer Prices.” In Dale W. Jorgenson and Ralph Landau, eds., Technology and Capital Formation. Cambridge, Mass.: MIT Press.

———. 1990. The Measurement of Durable Goods Prices. Chicago: University of Chicago Press.

———. 1999a. “U.S. Economic Growth since 1870: What We Know and Still Need to Know.” American Economic Review 89, no. 2 (May): 123-8. Papers and Proceedings of the 111th Annual Meeting of the American Economic Association.

———. 1999b. "Has the New Economy Rendered the Productivity Slowdown Obsolete?" Unpublished paper, Northwestern University.

———. 2000. "Does the 'New Economy' Measure up to the Great Inventions of the Past?" Journal of Economic Perspectives 14, no. 4 (fall): 49-74.

Gramlich, Edward M. 1990. "How Should Public Infrastructure Be Financed?" In Alicia H. Munnell, ed., Is There a Shortfall in Public Capital Investment? Federal Reserve Bank of Boston Conference Series no. 34: 222-37.

———. 1994. "Infrastructure Investments: A Review Essay." Journal of Economic Literature 32 (September): 1176-96.

Greenwood, Jeremy, Zvi Hercowitz, and Per Krusell. 1997. "Long-Run Implications of Investment-Specific Technological Change." American Economic Review 87, no. 3 (June): 342-62.

Griliches, Zvi. 1960. "Measuring Inputs in Agriculture: A Critical Survey." Journal of Farm Economics 42, no. 5 (December): 1411-27.

———. 1973. "Research Expenditures and Growth Accounting." In B. R. Williams, ed., Science and Technology in Economic Growth. London: Macmillan.

———. 1979. "Issues in Assessing the Contribution of Research and Development to Productivity Growth." Bell Journal of Economics 10, no. 1 (spring): 92-116.

———. 1992. "The Search for R&D Spillovers." Scandinavian Journal of Economics 94 (Supplement): 29-47.

———. 1994. "Productivity, R&D, and the Data Constraint." American Economic Review 84, no. 1 (March): 1-23.

———. 1995. "R&D and Productivity: Econometric Results and Measurement Issues." In P. Stoneman, ed., Handbook of the Economics of Innovation and Technological Change. Oxford: Basil Blackwell.

———. 1996. "Education, Human Capital, and Growth: A Personal Perspective." NBER Working Paper no. 5426.

Grossman, Gene M. 1996. Discussion of "Technology in Growth Theory." In Jeffrey C. Fuhrer and Jane Sneddon Little, eds., Technology and Growth. Federal Reserve Bank of Boston Conference Series no. 40: 83-9.

Grossman, Gene M., and Elhanan Helpman. 1991. Innovation and Growth in the Global Economy. Cambridge, Mass.: MIT Press.

Gullickson, William, and Michael J. Harper. 1999. "Possible Measurement Bias in Aggregate Productivity Growth." Monthly Labor Review 122, no. 2 (February): 47-67.

Hall, Bronwyn H. 1996. "The Private and Social Returns to Research and Development." In Bruce L. R. Smith and Claude E. Barfield, eds., Technology, R&D, and the Economy. Washington, D.C.: Brookings Institution and American Enterprise Institute.

Hall, Robert E., and Dale W. Jorgenson. 1967. "Tax Policy and Investment Behavior." American Economic Review 57, no. 3 (June): 391-414.

Hercowitz, Zvi. 1998. "The Embodiment Controversy: A Review Essay." Journal of Monetary Economics 41, no. 1: 217-24.

Ho, Mun Sing, and Dale W. Jorgenson. 1999. "The Quality of the U.S. Workforce, 1948-1995." Unpublished paper, Harvard University.

Hsieh, Chang-Tai. 1997. "What Explains the Industrial Revolution in East Asia? Evidence from Factor Markets." Unpublished paper, University of California at Berkeley, November.

———. 1999. "Productivity Growth and Factor Prices in East Asia." American Economic Review 89, no. 2 (May): 133-8. Papers and Proceedings of the 111th Annual Meeting of the American Economic Association.

Hulten, Charles R. 1979. "On the Importance of Productivity Change." American Economic Review 69, no. 1 (March): 126-36.

———. 1992. "Growth Accounting When Technical Change Is Embodied in Capital." American Economic Review 82, no. 4 (September): 964-80.

———. 2000. "Total Factor Productivity: A Short Biography." NBER Working Paper no. 7471, January.

Jones, Charles I. 1995a. "R&D-Based Models of Economic Growth." Journal of Political Economy 103 (August): 759-84.

———. 1995b. "Time Series Tests of Endogenous Growth Models." Quarterly Journal of Economics 110 (May): 495-525.

———. 1999. "Growth: With or without Scale Effects." American Economic Review 89, no. 2 (May): 139-44. Papers and Proceedings of the 111th Annual Meeting of the American Economic Association.

Jones, Charles I., and John C. Williams. 1998. "Measuring the Social Return to R&D." Quarterly Journal of Economics 63, no. 4 (November): 1119-36.

Jorgenson, Dale W. 1966. "The Embodiment Hypothesis." Journal of Political Economy 74, no. 1 (February): 1-17.

———. 1990. "Productivity and Economic Growth." In Ernst R. Berndt and Jack E. Triplett, eds., Fifty Years of Economic Measurement. Chicago: University of Chicago Press.

———. 1996. "Technology in Growth Theory." In Jeffrey C. Fuhrer and Jane Sneddon Little, eds., Technology and Growth. Federal Reserve Bank of Boston Conference Series no. 40: 45-77.

Jorgenson, Dale W., and Zvi Griliches. 1967. "The Explanation of Productivity Change." Review of Economic Studies 34, no. 99 (July): 249-80.

Jorgenson, Dale W., and Kevin J. Stiroh. 1999. "Information Technology and Growth." American Economic Review 89, no. 2 (May): 109-15. Papers and Proceedings of the 111th Annual Meeting of the American Economic Association.

———. 2000. "Raising the Speed Limit: U.S. Economic Growth in the Information Age." Brookings Papers on Economic Activity, no. 1: 125-211.

Keller, Wolfgang. 1998. "Are International R&D Spillovers Trade-Related? Analyzing Spillovers among Randomly Matched Trade Partners." European Economic Review 42, no. 8: 1469-81.

Kiley, Michael T. 1999. "Computers and Growth with Costs of Adjustment: Will the Future Look Like the Past?" Board of Governors of the Federal Reserve System Finance and Economics Discussion Series no. 36, July.

Kortum, Samuel. 1997. "Research, Patenting, and Technological Change." Econometrica 65, no. 6 (November): 1389-419.

Kremer, Michael. 1993. "Population Growth and Technological Change: One Million B.C. to 1990." Quarterly Journal of Economics 108, no. 3 (August): 681-716.

Krugman, Paul. 1994. "The Myth of Asia's Miracle." Foreign Affairs 73, no. 6 (November-December): 62-78.

Lehr, Bill, and Frank Lichtenberg. 1999. "Information Technology and Its Impact on Firm-Level Productivity: Evidence from Government and Private Data Sources, 1977-1993." Canadian Journal of Economics 32, no. 2: 335-62.

Lichtenberg, Frank R. 1995. "The Output Contribution of Computer Equipment and Personnel: A Firm-Level Analysis." Economics of Innovation and New Technology 3, nos. 3-4: 201-17.

Lucas, Robert E., Jr. 1988. "On the Mechanics of Economic Development." Journal of Monetary Economics 22, no 1: 3-42.

Maclean, Dinah. 1997. "Lagging Productivity Growth in the Service Sector: Mismeasurement, Mismanagement, or Misinfomation?" Bank of Canada Working Paper no. 97-6.

Mankiw, N. Gregory. 1995. "The Growth of Nations." Brookings Papers on Economic Activity, no. 1: 275-326.

Mankiw, N. Gregory, David Romer, and David Weil. 1992. "A Contribution to the Empirics of Economic Growth." Quarterly Journal of Economic Growth 107, no. 2 (May): 407-38.

McGuckin, Robert H., and Kevin J. Stiroh. 1998. "Computers, Productivity, and Growth." Conference Board Economic Research Report no. 1213-98-RR.

McGuckin, Robert H., Mary L. Streitwieser, and Mark Doms. 1998. "The Effect of Technology Use on Productivity Growth." Economics of Innovation and New Technology 7, no. 1: 1-26.

References (Continued)

Mincer, Jacob. 1958. "Investment on Human Capital and Personal Income Distribution." Journal of Political Economy 64, no. 4: 281-302.

———. 1974. Schooling, Experience, and Earnings. New York: Columbia University Press and National Bureau of Economic Research.

Morrison, Catherine J. 1997. "Assessing the Productivity of Information Technology Equipment in U.S. Manufacturing Industries." Review of Economics and Statistics 79 (August): 471-81.

Munnell, Alicia H., ed. 1990. Is There a Shortfall in Public Capital Investment? Federal Reserve Bank of Boston Conference Series no. 34.

Oliner, Stephen D., and Daniel E. Sichel. 2000. "The Resurgence of Growth in the Late 1990s: Is Information Technology the Story?" Journal of Economic Perspectives 14, no. 4 (fall): 3-22.

Parry, Robert T. 2000. "Implications of Productivity Uncertainty for Monetary Policy." Business Economics 35, no. 1 (January): 13-5.

Rodrick, Dani. 1997. "TFPG Controversies, Institutions, and Economic Performance in East Asia." NBER Working Paper no. 5914, February.

Romer, Paul. 1986. "Increasing Returns and Long-Run Growth." Journal of Political Economy 94, no. 5 (October): 1002-37.

———. 1990. "Endogenous Technological Change." Journal of Political Economy 98, no. 5 (part 2): S71-101.

———. 1994. "The Origins of Endogenous Growth." Journal of Economic Perspectives 8, no. 1 (winter): 3-22.

Schultz, Theodore W. 1961. "Investment in Human Capital." American Economic Review 51, no. 1 (March): 1-17.

Segerstrom, Paul S. 1998. "Endogenous Growth without Scale Effects." American Economic Review 88, no. 5 (December): 1290-310.

Sichel, Daniel E. 1997. The Computer Revolution: An Economic Perspective. Washington, D.C.: Brookings Institution.

Siegel, Donald. 1997. "The Impact of Computers on Manufacturing Productivity Growth: A Multiple Indicators, Multiple Causes Approach." Review of Economics and Statistics 1 (February): 68-78.

Siegel, Donald, and Zvi Griliches. 1992. "Purchased Services, Outsourcing, Computers, and Productivity in Manufacturing." In Zvi Griliches, ed., Output Measurement in the Service Sector. Chicago: University of Chicago Press.

Solow, Robert M. 1956. "A Contribution to the Theory of Economic Growth." Quarterly Journal of Economics 70, no. 1 (February): 56-94.

———. 1957. "Technical Change and the Aggregate Production Function." Review of Economics and Statistics 39, no. 3: 313-30.

———. 1960. "Investment and Technical Progress." In Kenneth J. Arrow, Samuel Karlin, and Patrick Suppes, eds., Mathematical Methods in the Social Sciences. Stanford, Calif.: Stanford University Press.

Steindel, Charles. 1992. "Manufacturing Productivity and High-Tech Investment." Federal Reserve Bank of New York Quarterly Review, summer: 39-47.

Stiroh, Kevin J. 1998a. "Computers, Productivity, and Input Substitution." Economic Inquiry 36, no. 2 (April): 175-91.

———. 1998b. "Long-Run Growth Projections and the Aggregate Production Function: A Survey of Models Used by the U.S. Government." Contemporary Economic Policy 16 (October): 467-79.

Triplett, Jack E. 1989. "Price and Technological Change in a Capital Good: A Survey of Research on Computers." In Dale W. Jorgenson and Ralph Landau, eds., Technology and Capital Formation. Cambridge, Mass.: MIT Press.

———. 1996. "High-Tech Industry Productivity and Hedonic Price Indices." OECD Proceedings: Industry Productivity, International Comparison, and Measurement Issues, 119-42.

References (Continued)

———. 1999a. "The Solow Productivity Paradox: What Do Computers Do to Productivity?" Canadian Journal of Economics 32, no. 2 (April): 309-34.

———. 1999b. "Economic Statistics, the New Economy, and the Productivity Slowdown." Business Economics 34, no. 2 (April): 13-7.

Whelan, Karl. 2000. "Computers, Obsolescence, and Productivity." Board of Governors of the Federal Reserve System Finance and Economics Discussion Series no. 6.

Wolff, Edward N. 1996. "The Productivity Slowdown: The Culprit at Last? Follow-Up on Hulten and Wolff." American Economic Review 86, no. 5: 1239-52.

Young, Alwyn. 1995. "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience." Journal of Quarterly Economics 110, no. 3 (August): 641-80.

———. 1998a. "Growth without Scale Effects." Journal of Political Economy 106, no. 1: 41-63.

———. 1998b. "Alternative Estimates of Productivity Growth in the NICs: A Comment on the Findings of Chang-Tai Hsieh." NBER Working Paper no. 6657.

Ray C. Fair

Actual Federal Reserve Policy Behavior and Interest Rate Rules

- Several studies have attempted to model how the Federal Reserve makes policy choices affecting interest rates. These studies have yielded “rules” relating the interest rate that the Fed is assumed to control to a set of variables thought to affect Fed policy behavior.

- Many of these studies conclude that interest rate rules do not have stable coefficient estimates over time—a finding that suggests structural change in the Fed’s policy behavior.

- A specification of an interest rate rule, estimated over the 1954:1–1979:3 and the 1982:4–1999:3 periods, does pass a stability test.

- Nevertheless, the results show a large economic difference in the coefficient on inflation between the first and second periods, and the relatively restrained behavior of inflation in recent years makes it hard to determine whether there has been a structural break.

Economists attempting to approximate the policy behavior of the Federal Reserve often have done so by estimating interest rate rules for the United States. In these rules, the variable on the left-hand side is the interest rate that the Federal Reserve is assumed to control, while the variables on the right-hand side are those that are assumed to affect Federal Reserve behavior.

There have been numerous examples of estimated interest rate rules over the past forty years. The first appeared in Dewald and Johnson (1963), who regressed the Treasury bill rate on a constant, the Treasury bill rate lagged once, real GNP, the unemployment rate, the balance-of-payments deficit, and the consumer price index. The next example can be found in Christian (1968), followed by many others. In 1978, I added an estimated interest rate rule to my U.S. macroeconometric model (Fair 1978). Later, McNees (1986, 1992) estimated rules in which some of the explanatory variables were the Federal Reserve’s internal forecasts of different variables, and Khoury (1990) provided an extensive list of estimated rules through 1986. More recently, Judd and Rudebusch (1998) estimated rules for various subsets of the 1970-97 period and Clarida, Galí, and Gertler (2000) estimated them for different Federal Reserve chairmen.

An interesting question that arises from these studies is whether the policy behavior of the Federal Reserve has changed over time. If one interprets such behavior as being approximated by a particular rule, this question can then be viewed as whether the coefficients in the rule have changed

Ray C. Fair is the John M. Musser Professor of Economics at Yale University.

over time. There seems to be a general view in the recent literature that estimated interest rate rules do not have stable coefficient estimates over time. For example, Judd and Rudebusch (1998, p. 3) observe that, "Overall, it appears that there have not been any great successes in modeling Federal Reserve behavior with a single, stable reaction function."

It seems clear that the Federal Reserve's policy behavior over the 1979:4-1982:3 period (which I refer to as the "early Volcker" period) differed from that of other periods.[1] The stated policy of the Federal Reserve during this period was to focus more on monetary aggregates than it had done before. Any interesting stability question must therefore exclude this period, since any hypothesis of stability that includes it is quite likely to be rejected. One obvious hypothesis to test is whether a rule's coefficients were the same before 1979:4 as they were after 1982:3. In a recent paper (Fair 2000), I tested this hypothesis using the 1978 specified rule mentioned above, and it was not rejected. Further test results are presented in the next section of this article.

If one finds a rule that seems to be a good approximation of the Federal Reserve's policy behavior, how should the residuals from the equation be interpreted? It is important to realize that estimated interest rate rules in general are not optimal. If the Federal Reserve behaves by minimizing the expected value of a loss function subject to a model of the economy, its optimal rule depends on all the predetermined variables in the model. The coefficients in the optimal rule are a combination of the coefficients in the loss function and the coefficients in the model. In this case, any estimated rule is just an approximation of the optimal rule, where the approximation may or may not be any good. If we assume that the Federal Reserve optimizes, then the actual values of the interest rate are the optimal values, so a residual for an estimated rule is the difference between the predicted value from the rule and the optimal (actual) value. The residuals are "mistakes" made by the econometrician, not by the Federal Reserve. This line of reasoning is pursued later on.

> *It seems clear that the Federal Reserve's policy behavior over the 1979:4-1982:3 period . . . differed from that of other periods. The stated policy of the Federal Reserve during this period was to focus more on monetary aggregates than it had done before.*

Estimated Interest Rate Rules

The rule that I added to my U.S. macroeconometric model in 1978 has been changed slightly over time. The main modification has been the addition of a dummy variable term to account for the change in Federal Reserve operating procedure during the early Volcker period. As noted above, the Federal Reserve's stated policy during this period was to focus more on monetary aggregates than it had done in the past. The estimated interest rate rule already had the lagged growth of the money supply as an explanatory variable, and the change in policy was modeled by adding the lagged growth of the money supply multiplied by a dummy variable as another explanatory variable. The dummy variable is 1 for the 1979:4-1982:3 period and 0 otherwise.

The specification of the rule used in this article is

$$(1)\quad r = \alpha_1 + \alpha_2 \dot{p} + \alpha_3 u + \alpha_4 \Delta u + \alpha_5 \dot{m}_{-1} + \alpha_6 D1 \times \dot{m}_{-1} + \alpha_7 r_{-1} + \alpha_8 \Delta r_{-1} + \alpha_9 \Delta r_{-2} + \varepsilon,$$

where r is the three-month Treasury bill rate, $\dot{p}$ is the quarterly rate of inflation at an annual rate, u is the unemployment rate, $\dot{m}$ is the quarterly rate of growth of the money supply at an annual rate, and $D1$ equals 1 for 1979:4-1982:3 and 0 otherwise. The estimates of equation 1 for three different sample periods are presented in Table 1.[2]

The endogenous variables on the right-hand side of equation 1 are inflation and the unemployment rate, and the two-stage least squares technique is used to estimate the equation. In the first-stage regression, inflation and the unemployment rate are regressed on a set of predetermined variables (the main variables in the U.S. model, which can be found at <http://fairmodel.econ.yale.edu>). The predicted values from these regressions are then used in the second stage. One can look at these regressions as those used by the Federal Reserve to predict inflation and the unemployment rate, so one need not assume that the Federal Reserve has perfect foresight.

If the Federal Reserve's expectations of *future* values of inflation and the unemployment rate affect its current decision, these expectations should be added to equation 1. A way to test this possibility is to add future values of inflation and the unemployment rate to equation 1 and then estimate it using Hansen's (1982) method-of-moments estimator, where the instruments used are the main predetermined variables in the U.S. model. Hansen's method in this context is simply two-stage least squares adjusted to account for the serial correlation properties of the error term. The test is to see if the future values are statistically significant. I have performed this test on various versions of my estimated interest rate rules using different lead lengths, and the lead values do not turn out to be significant.[3] Thus, there is no evidence that future values are needed in equation 1, and they have not been used. Clarida, Galí, and Gertler (2000) use future values in many of their specifications, but they point out (p. 164) that their conclusions are not changed if they do not use these values.

Equation 1 is a "leaning-against-the-wind equation." The variable *r* is estimated to depend positively on the inflation rate and the lagged growth of the money supply and negatively on the unemployment rate and the change in the unemployment rate. Adjustment and smoothing effects are captured by the lagged values of *r*. The coefficient on the lagged money supply growth is more than ten times larger for the early Volcker period than it is for the period before or after—a finding that is consistent with the Federal Reserve's stated policy of focusing more on monetary aggregates during this period. This method of accounting for the Federal Reserve's policy shift does not, of course, capture the richness of the change in behavior, but at least it seems to capture some of the change.

The Wald value in Table 1 is used to test the hypothesis that the coefficients in the 1954:1-1979:3 period are the same as those in the 1982:4-1999:3 period. (The early Volcker period is excluded from this test, so the *D*1 term is excluded.) The Wald statistic is presented in Andrews and Fair (1988, equation 3.6). It has the advantage of working under very general assumptions about the properties of the error terms and can be used when the estimator is two-stage least squares, which it is here. The Wald statistic is distributed as x^2, with (in the present case) eight degrees of freedom. The estimates of the equation for the two subperiods are presented in Table 1. The value of the Wald statistic is 11.13, which has a *p*-value of .194. The hypothesis of equality thus is not rejected even at the 10 percent level.

Equation 1, estimated for the entire 1954:1-1999 period, underwent a number of other tests.[4] First, the lagged values of all the variables in the equation $(r_{-4}, \dot{p}_{-1}, u_{-2}, \dot{m}_{-2}, D1 \times \dot{m}_{-2})$ were added and the joint significance of these variables was tested. The x^2 value was 5.69, with five degrees of freedom, which has a *p*-value of .338. Adding these variables encompasses a number of alternative hypotheses about the dynamics, and these hypotheses are rejected because the added variables are not significant.[5] Second, the equation was estimated assuming first-order serial correlation of the error term. The x^2 value was 1.30, with one degree of freedom, which has a *p*-value of .255. Third, the percentage change in

Table 1
Estimated U.S. Interest Rate Rule

	1954:1-1999:3		1954:1-1979:3		1982:4-1999:3		1954:1-1999:3	
	Coefficient	*t*-Statistic	Coefficient	*t*-Statistic	Coefficient	*t*-Statistic	Coefficient	*t*-Statistic
Constant	.855	5.42	.762	3.36	.409	1.93	.663	3.98
$\dot{p}$	.071	4.17	.077	3.05	.145	3.30	.080	4.35
u	-.131	-4.18	-.114	-3.00	-.085	-2.29	-.111	-3.55
Δu	-.748	-6.05	-.380	-3.04	-.901	-4.53	-.567	-4.43
$\dot{m} - 1$	.014	2.26	.027	3.65	.001	0.16	.010	1.73
$D1 \times \dot{m} - 1$	.219	9.71	-		-		.349	8.72
$r - 1$	.916	47.32	.887	21.60	.939	36.17	.922	38.71
$\Delta r - 1$	.210	3.75	.251	2.89	.280	2.88	.338	5.36
$\Delta r - 2$	-.345	-6.71	-.225	-2.54	-.195	-2.26	-.357	-7.23
$D1 \times \dot{p}$							-.148	-3.40
$D2 \times \dot{p}$							.060	1.67
SE	.471		.411		.317		.450	
R^2	.971		.960		.970		.974	
DW	1.85		1.82		2.09		2.17	
Wald (*p*-value)	11.13 (.194)							
Number of observations	183		103		68		183	

Source: Author's calculations.

Notes: The dependent variable is *r*. The estimation period is 1954:1-1999:3. The estimation technique is two-stage least squares. *r* is the three-month Treasury bill rate, $\dot{p}$ is the inflation rate, *u* is the unemployment rate, and $\dot{m}$ is the growth rate of the money supply. *D*1 equals 1 for 1979:4-1982:3, 0 otherwise; *D*2 equals 1 for 1982:4-1999:3, 0 otherwise.

real GDP was added (without excluding the change in the unemployment rate). The x^2 value was 0.51, with one degree of freedom, which has a p-value of .476. Finally, an output gap variable and the change in the variable were added (without excluding the unemployment rate and the change in the unemployment rate).[6] The x^2 value was 2.43, with two degrees of freedom, which has a p-value of .297. Overall, the equation performs well in these tests. The added variables, including the output gap and the change in the output gap, do not have additional explanatory power.

When the unemployment rate and the change in the unemployment rate are added to the equation with the output gap and the change in the output gap already included, the x^2 value is 9.53, with two degrees of freedom, which has a p-value of .009. The unemployment rate and the change in the unemployment rate thus have additional explanatory power.

One likely reason why the stability hypothesis generally has been rejected is that most tests have included the [1979:4–1982:3] period, which is clearly different from the periods before and after.

Since the output gap and the change in the output gap do not have such power, in this sense the unemployment rate dominates the output gap. Many interest rate rules in the literature use some measure of the output gap as an explanatory variable, and the current results suggest that the unemployment rate may be a better variable.

Returning to the stability test, again note that the passing of this test is contrary to the overall view in the literature. One likely reason why the stability hypothesis generally has been rejected is that most tests have included the early Volcker period, which is clearly different from the periods before and after. The tests in Judd and Rudebusch (1998), for example, include the early Volcker period.

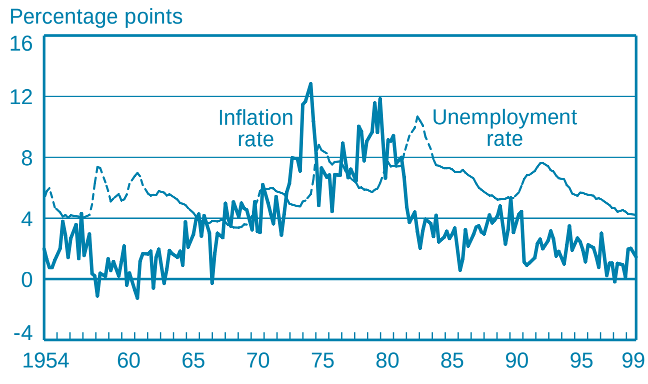
Clarida, Galí, and Gertler (2000) do not perform any stability tests; they simply note that the coefficient estimates for the different periods look quite different, especially the inflation coefficient. The equations for the two subperiods in Table 1 also show a large difference in the inflation coefficient. For the first subperiod, the long-run coefficient is 0.68 [=.077/(1.0 - .887)], and for the second subperiod it is 2.38 [=.145/(1.0 - .939)]. The Clarida, Galí, and Gertler coefficients (p. 150) are .83 for their pre-Volcker period (1960:1-1979:2) and 2.15 for their Volcker-Greenspan period (1979:3-1996:4).

Although the inflation coefficients seem quite different in Table 1, the Wald test does not reject the hypothesis of stability. It could be, however, that the test had low power, so another test was performed. This test is represented in the last two columns of Table 1. It is based on the assumption that all the coefficients are constant across time, except for the inflation coefficient, which is postulated to be different in each of the three subperiods: 1954:1-1979:3, 1979:4-1982:3, and 1982:4-1999:3, which I refer to as the "first," the "early Volcker," and the "second" periods, respectively. The coefficient estimate for $D1 \times \dot{p}$ is the estimated difference between the early Volcker period and the first period. This difference is not of much interest, however, since the added variable is just meant to dummy out the early Volcker period. The estimated difference is negative and significant (with a t-statistic of -3.40). The total coefficient for this period is -0.068 [= .080 - .148]. This negative value is not sensible, reflecting the fact that the early Volcker period is unusual and hard to model. (This is why the period was completely ignored in the Wald test.)

The coefficient estimate for $D2 \times \dot{p}$ is the estimated difference between the second and first periods. This estimated difference is .060, with a t-statistic of 1.67, which is not significant at the 5 percent level for a two-tailed test. Again, the long-run inflation coefficient for the second period of 1.79 [= .080 + .060)/(1 - .922)] is noticeably larger than that for the first period of 1.03 [= .080/(1 - .922)].

The results thus show a large economic, but not statistically significant, difference for the inflation coefficient between the first and second periods. One important fact to keep in mind is that the variance of inflation is much smaller in the second period than it is in the first. This can be seen from Chart 1, where inflation and the unemployment rate are plotted for the

Chart 1
Unemployment and Inflation: 1954:1-1999:3



Source: U.S. Department of Labor, Bureau of Labor Statistics.

1954:1-1999:3 period. The largest value of inflation in the second period is 5.33 percent, in 1990:1, and no other value is above 5 percent. However, the largest value for the first period is 12.83 percent, in 1974:3, and twenty-nine other values are above 5 percent.

If inflation were to rise substantially in the future, it would make for an interesting test of whether there has been a structural change in Federal Reserve policy behavior. The third equation in Table 1 implies a much larger Federal Reserve response than the first equation does; the test is which equation predicts the actual Federal Reserve response better. If the third equation is the better predictor, we will have strong evidence in favor of a shift in behavior from the earlier period. If the first equation excels, it will suggest that focusing only on the post-1982 period, when inflation has been low, has resulted in misleading estimates (in effect, a small-sample problem). In short, although the statistical tests in this article suggest that there has not been a shift in behavior, more observations are needed—particularly high-inflation ones—before much confidence can be placed in any conclusion.

Although the statistical tests in this article suggest that there has not been a shift in behavior, more observations are needed . . . before much confidence can be placed in any conclusion.

Deviations from the Rule

Equation 1, estimated for the 1954:1-1999:3 period, was solved dynamically for this period using the actual values of inflation, the unemployment rate, and the growth of the money supply. Chart 2 plots the predicted values from this simulation along with the actual values, and the appendix table presents the values. The actual values of inflation, the unemployment rate, and the growth of the money supply are also presented in the appendix table. For this exercise, the Federal Reserve is assumed to know the current values of inflation and the unemployment rate, since the actual values of these two variables are used.

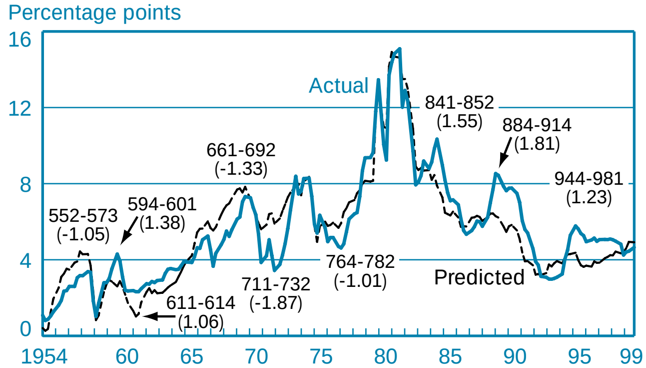
Nine subperiods appear in Chart 2. They represent periods in which the actual values differ from the predicted values by noticeable amounts for a number of consecutive quarters. There are six quarters in the early 1960s in which the interest rate was noticeably higher than predicted: 1959:4, 1960:1, and 1961:1-1961:4, and there are three subperiods from the mid-1980s for which this was true: 1984:1-1985:2, 1988:4-1991:4, and 1994:4-1998:1. Conversely, the interest rate was noticeably lower than predicted in the mid-1950s, in the late 1960s, and in two periods in the 1970s: 1955:2-1957:3, 1966:1-1969:2, 1971:1-1973:2, and 1976:4-1978:2. Chart 2 presents the average deviation for each of these subperiods in parentheses. The largest average deviation in absolute value is -1.87 percentage points, for the 1971:1-1973:2 period.

If the Federal Reserve is behaving optimally, then the deviations [from the interest rate rule] are actually errors made by the econometrician.

Taylor (1999) presents charts similar to Chart 2 using calibrated interest rate rules, and he interprets the deviations as policy mistakes. According to this interpretation, Chart 2 shows that the Federal Reserve's policy was too tight in the early 1960s, too loose in the late 1960s and in about half of the 1970s, and too tight in the mid-1980s, the late 1980s to early 1990s, and part of the second half of the 1990s.

However, if the Federal Reserve is behaving optimally, then the deviations are actually errors made by the econometrician. At each Federal Open Market Committee meeting, the Federal Reserve clearly knows more than is reflected in the rule, even if

Chart 2
Actual and Predicted Values of *r*: 1954:1-1999:3



Sources: Board of Governors of the Federal Reserve System; author's calculations.

Notes: Equation 1 in the text was estimated for the 1954:1-1999:3 period. The average deviations are in parentheses.

it is not formally solving an optimal control problem. According to this interpretation, deviations from the rule can be viewed as the use of more information by the Federal Reserve.

The Stabilization Effectiveness of Rules

There is a large literature examining the stabilization effectiveness of different interest rate rules. The general approach in this literature is to choose a rule and then use a model of the economy to examine how the economy would have behaved under the rule.[7] Using a calibrated model of the economy, Clarida, Galí, and Gertler (2000) show that interest rate rules with inflation coefficients less than 1 can be destabilizing. They then criticize the Federal Reserve's pre-1979 policy for being too timid, but praise it after 1979. This evaluation of policy is based on the assumption that there was a change in behavior after 1979: the Federal Reserve used a coefficient less than 1 before that year and a coefficient greater than 1 after it.

> *The judging of interest rate rules . . . can be sensitive to the economic model used.*

A different conclusion about interest rate rules is reached in Fair (2000) using the U.S. model mentioned above. Stochastic simulation was used to compute variances of the endogenous variables for different rules.[8] The simulation period was 1993:1-1998:4.

The estimated rule in this article—estimated for the 1954:1-1999:2 period—was first tried. The variances using this rule for real output, the price level, and the Treasury bill rate are presented in the first row of Table 2. Two calibrated rules were then tried. The first was the Taylor rule, which has a coefficient of 0.5 on the output gap and 1.5 on inflation. The second was a rule with a coefficient of 0.5 on the output gap but only 0.25 on inflation; this will be called the ".25 rule." The variances using these two rules are presented in the second and third rows of Table 2, respectively.

Compared with the estimated and .25 rules, the Taylor rule achieved a lower price level variance (0.61, versus 0.69 and 0.71) at the cost of a considerably higher interest rate variance (2.86, versus 1.14 and 1.19). Some insight into this result can be gleaned from a property of the price equation in the U.S. model, which is that the price level responds only modestly to demand (a common feature of most estimated price equations). Since the interest rate affects the price level primarily through its effects on demand, the price level responds only modestly to interest rate changes. The Taylor rule has a large coefficient on inflation, so a large price shock leads to a large change in the interest rate, but this in turn has only a modest impact on offsetting the effects of the price shock. For a rule like the estimated or the .25 rule, the interest rate responds much less to a price shock, so the interest rate variance is smaller. The cost of a smaller interest rate response in terms of offsetting the effects of the price shock is modest because of the modest effect of the interest rate on the price level.

Why are the estimated and .25 rules not destabilizing, as they would be in the Clarida, Galí, and Gertler model? The answer is that the response of output to a price shock is much different in that model than it is in the U.S. model. Consider a positive price shock with no change in the nominal interest rate. In the Clarida, Galí, and Gertler model, this shock is expansionary because the real interest rate, which has a negative effect on output, is lower. In the U.S. model, however, a positive price shock with no change in the nominal interest rate is contractionary. In the short run, the aggregate price level rises more than wage rates rise, so a fall in real income occurs. Real wealth also falls. These effects are contractionary on demand. In addition, the empirical results suggest that households respond to nominal interest rates and not real interest rates, so there is no positive household response to lower real interest rates. The net effect of a positive price shock with no change in the nominal interest rate is contractionary in the U.S. model. If this is true, then the Federal Reserve, in response to a positive price shock, does not have to increase the nominal interest rate more than the increase in inflation to

Table 2
Variability Estimates

	Y	*P*	*r*
Estimated rule	4.12	0.69	1.14
Calibrated rule (1.50, Taylor)	4.04	0.61	2.86
Calibrated rule (.25)	3.57	0.71	1.19

Source: Author's calculations.
Notes: The simulation period is 1993:1-1998:4. Y is real GDP, P is the GDP deflator, and r is the three-month Treasury bill rate. The variability measures are the squares of percentage points. The rules are described in the text.

achieve a contraction. There will be a contraction even if there is no increase in the nominal interest rate at all!

The judging of interest rate rules therefore can be sensitive to the economic model used. Using an economic model in which positive price shocks are expansionary—as Clarida, Galí, and Gertler do—leads to a much different conclusion than using a macroeconometric model like the U.S. model, in which positive price shocks are contractionary. Using small, calibrated models to make policy conclusions is risky if the models are at odds with more empirically based ones. It may be that the specification and calibration have not captured reality well.

Appendix: Variable Values, 1954:1-1999:3

Quarter	r	$\hat{r}$	$r-\hat{r}$	$\dot{p}$	u	$\dot{m}$
1954:1	1.08	0.41	0.67	1.99	5.23	0.85
1954:2	0.81	0.27	0.55	1.29	5.78	3.69
1954:3	0.87	0.36	0.51	0.75	5.97	10.01
1954:4	1.04	1.23	-0.19	0.75	5.36	-1.09
1955:1	1.26	1.94	-0.68	1.23	4.71	6.45
1955:2	1.61	2.52	-0.91	2.00	4.38	0.17
1955:3	1.86	3.10	-1.24	3.79	4.11	5.58
1955:4	2.35	3.27	-0.93	2.89	4.21	-3.17
1956:1	2.38	3.53	-1.15	1.42	4.03	2.33
1956:2	2.60	3.48	-0.89	2.67	4.19	0.92
1956:3	2.60	3.85	-1.25	3.58	4.13	1.76
1956:4	3.06	3.91	-0.85	1.34	4.10	0.27
1957:1	3.17	4.44	-1.27	4.30	3.95	1.29
1957:2	3.16	4.29	-1.14	1.55	4.06	-0.04
1957:3	3.38	4.30	-0.91	2.96	4.21	1.14
1957:4	3.34	3.70	-0.36	0.35	4.92	-4.16
1958:1	1.84	2.28	-0.45	0.20	6.28	3.48
1958:2	1.02	0.83	0.18	-1.11	7.36	0.06
1958:3	1.71	1.07	0.64	0.38	7.31	10.24
1958:4	2.79	2.30	0.49	0.18	6.35	2.60
1959:1	2.80	2.73	0.07	1.33	5.80	4.84
1959:2	3.02	2.94	0.08	0.56	5.10	0.75
1959:3	3.53	2.85	0.68	1.15	5.29	3.41
1959:4	4.30	2.60	1.70	0.20	5.59	-4.99
1960:1	3.94	2.88	1.06	1.15	5.16	0.58
1960:2	3.09	2.58	0.51	2.17	5.23	-0.27
1960:3	2.39	2.16	0.23	-0.40	5.55	5.29
1960:4	2.36	1.73	0.63	0.40	6.25	-2.63
1961:1	2.38	1.31	1.06	-0.72	6.77	7.69
1961:2	2.32	1.02	1.30	-1.25	6.97	1.10
1961:3	2.32	1.15	1.17	1.18	6.75	6.72
1961:4	2.48	1.76	0.71	1.67	6.17	0.74
1962:1	2.74	2.31	0.43	1.65	5.61	1.94
1962:2	2.72	2.51	0.21	1.83	5.48	0.66
1962:3	2.86	2.25	0.61	-0.59	5.54	3.09
1962:4	2.80	2.40	0.40	1.44	5.51	4.82
1963:1	2.91	2.24	0.67	1.96	5.78	1.22
1963:2	2.94	2.27	0.67	-0.28	5.68	2.46
1963:3	3.28	2.40	0.88	0.63	5.49	6.50
1963:4	3.50	2.54	0.95	1.88	5.57	3.77
1964:1	3.54	2.65	0.88	1.67	5.46	1.58
1964:2	3.48	2.83	0.65	1.43	5.22	2.70
1964:3	3.50	3.11	0.39	1.83	4.99	15.32
1964:4	3.69	3.38	0.31	0.91	4.95	1.87
1965:1	3.90	3.69	0.21	3.76	4.87	0.21
1965:2	3.88	3.92	-0.04	2.10	4.66	-1.29
1965:3	3.86	4.22	-0.36	2.94	4.35	14.91
1965:4	4.16	4.85	-0.69	3.94	4.10	6.35
1966:1	4.63	5.44	-0.81	4.27	3.85	4.27
1966:2	4.60	5.63	-1.03	2.82	3.81	-3.27
1966:3	5.05	5.64	-0.60	4.17	3.75	7.59
1966:4	5.25	6.02	-0.77	3.03	3.68	2.31
1967:1	4.53	5.67	-1.14	-0.27	3.81	6.10
1967:2	3.66	5.54	-1.88	1.68	3.81	-0.54
1967:3	4.34	5.72	-1.37	3.01	3.78	18.80
1967:4	4.79	6.37	-1.59	2.72	3.92	6.92
1968:1	5.06	6.65	-1.59	4.98	3.73	3.60
1968:2	5.51	6.86	-1.35	3.97	3.54	0.89
1968:3	5.23	6.96	-1.73	3.48	3.51	15.26
1968:4	5.58	7.23	-1.65	5.07	3.39	13.59
1969:1	6.14	7.69	-1.55	4.09	3.38	-1.31
1969:2	6.24	7.75	-1.51	4.99	3.42	1.15
1969:3	7.05	7.53	-0.49	4.64	3.59	6.25
1969:4	7.32	7.83	-0.52	4.54	3.58	7.88
1970:1	7.26	7.17	0.09	3.23	4.16	9.29
1970:2	6.75	6.74	0.01	5.08	4.75	5.13
1970:3	6.37	6.24	0.13	3.13	5.17	21.44
1970:4	5.36	5.94	-0.59	3.07	5.80	-14.35
1971:1	3.86	5.60	-1.74	6.21	5.91	16.99
1971:2	4.21	5.84	-1.63	4.99	5.91	10.58
1971:3	5.05	6.40	-1.35	4.27	5.98	12.58
1971:4	4.23	6.44	-2.21	3.63	5.95	-12.19
1972:1	3.44	5.90	-2.46	5.42	5.77	18.84
1972:2	3.75	6.17	-2.42	2.89	5.66	6.14
1972:3	4.24	6.56	-2.32	4.25	5.58	10.20
1972:4	4.85	6.91	-2.06	5.71	5.30	5.01
1973:1	5.64	7.28	-1.64	6.29	4.95	5.95
1973:2	6.61	7.53	-0.92	7.96	4.89	12.45
1973:3	8.39	7.86	0.53	7.90	4.79	-0.09
1973:4	7.46	7.99	-0.52	7.10	4.77	8.46
1974:1	7.60	7.39	0.21	11.47	5.09	3.25
1974:2	8.27	8.11	0.16	11.67	5.16	8.56
1974:3	8.29	8.36	-0.07	12.83	5.58	-0.63
1974:4	7.34	7.42	-0.09	10.37	6.56	4.82
1975:1	5.87	5.78	0.09	8.26	8.22	3.38
1975:2	5.40	4.94	0.46	4.83	8.83	20.43
1975:3	6.34	5.74	0.59	7.30	8.47	-1.66
1975:4	5.68	6.00	-0.32	6.69	8.26	1.03
1976:1	4.95	5.78	-0.82	6.84	7.72	10.21
1976:2	5.17	5.83	-0.66	4.44	7.53	7.94
1976:3	5.17	5.95	-0.78	6.87	7.70	0.20
1976:4	4.70	5.68	-0.98	6.81	7.73	9.49
1977:1	4.62	5.92	-1.29	8.93	7.49	12.65
1977:2	4.83	6.50	-1.67	7.76	7.10	6.01
1977:3	5.47	6.71	-1.24	6.66	6.86	9.01
1977:4	6.14	7.02	-0.89	7.22	6.61	6.38
1978:1	6.41	7.14	-0.73	6.46	6.33	7.82
1978:2	6.48	7.50	-1.02	10.04	6.00	9.66
1978:3	7.32	7.66	-0.34	9.69	6.02	7.52
1978:4	8.68	8.00	0.68	7.77	5.88	6.20
1979:1	9.36	8.15	1.21	9.05	5.88	4.53
1979:2	9.37	8.11	1.26	9.66	5.71	11.85
1979:3	9.63	8.15	1.48	11.57	5.87	14.40
1979:4	11.80	11.57	0.24	9.64	5.94	5.69

Source: Author's calculations.

Appendix: Variable Values, 1954:1-1999:3 (Continued)

Quarter	r	$\hat{r}$	$r-\hat{r}$	$\dot{p}$	u	$\dot{m}$
1980:1	13.46	12.89	0.57	11.85	6.30	-0.09
1980:2	10.05	10.98	-0.93	6.63	7.32	8.29
1980:3	9.24	10.93	-1.69	9.14	7.68	10.40
1980:4	13.71	14.17	-0.46	9.07	7.40	0.82
1981:1	14.37	14.92	-0.55	9.42	7.43	8.48
1981:2	14.83	14.68	0.15	7.59	7.40	3.71
1981:3	15.09	14.62	0.47	8.00	7.42	2.47
1981:4	12.02	13.49	-1.47	6.70	8.24	9.85
1982:1	12.90	13.50	-0.60	4.72	8.84	-0.25
1982:2	12.36	12.99	-0.63	3.73	9.43	-0.39
1982:3	9.71	10.88	-1.17	4.40	9.94	13.74
1982:4	7.94	8.90	-0.97	3.10	10.68	13.26
1983:1	8.08	8.72	-0.64	2.03	10.40	15.76
1983:2	8.42	8.82	-0.40	3.17	10.10	9.50
1983:3	9.19	8.69	0.49	3.93	9.36	7.39
1983:4	8.79	8.71	0.08	3.65	8.54	3.38
1984:1	9.13	8.20	0.93	2.79	7.87	19.73
1984:2	9.84	8.45	1.39	4.19	7.48	3.06
1984:3	10.34	7.89	2.45	2.44	7.45	-2.33
1984:4	8.97	7.28	1.70	2.75	7.28	10.01
1985:1	8.18	6.47	1.72	3.14	7.28	6.49
1985:2	7.52	6.40	1.12	2.66	7.29	11.81
1985:3	7.10	6.33	0.77	2.94	7.21	22.83
1985:4	7.15	6.53	0.61	3.35	7.05	5.61
1986:1	6.89	6.22	0.67	0.58	7.02	16.02
1986:2	6.13	5.73	0.40	1.33	7.18	19.97
1986:3	5.53	5.76	-0.23	3.24	6.99	7.98
1986:4	5.34	5.75	-0.41	2.17	6.84	31.13
1987:1	5.53	6.21	-0.67	2.91	6.62	-6.63
1987:2	5.73	6.23	-0.50	3.41	6.28	8.51
1987:3	6.03	6.31	-0.28	3.50	6.01	4.41
1987:4	6.00	6.25	-0.25	3.09	5.87	0.36
1988:1	5.76	6.04	-0.28	2.96	5.73	9.72
1988:2	6.23	6.24	-0.01	4.21	5.49	10.22
1988:3	6.99	6.43	0.56	3.69	5.49	0.46
1988:4	7.70	6.43	1.27	3.88	5.35	-3.86
1989:1	8.53	6.29	2.24	4.14	5.22	2.99
1989:2	8.44	6.22	2.22	4.80	5.24	3.42
1989:3	7.85	5.74	2.11	2.30	5.28	-0.61
1989:4	7.63	5.46	2.17	3.25	5.37	0.03

Quarter	r	$\hat{r}$	$r-\hat{r}$	$\dot{p}$	u	$\dot{m}$
1990:1	7.76	5.76	2.00	5.33	5.30	7.95
1990:2	7.77	5.83	1.94	3.06	5.34	7.30
1990:3	7.49	5.54	1.95	4.25	5.69	1.88
1990:4	7.02	5.09	1.93	4.44	6.11	-0.32
1991:1	6.05	4.38	1.67	1.12	6.57	8.13
1991:2	5.59	3.92	1.67	0.91	6.82	6.62
1991:3	5.41	3.93	1.48	1.06	6.85	10.89
1991:4	4.58	3.71	0.88	1.40	7.10	6.76
1992:1	3.91	3.22	0.69	2.34	7.38	20.21
1992:2	3.72	3.25	0.48	2.62	7.60	4.67
1992:3	3.13	3.20	-0.07	1.99	7.63	14.48
1992:4	3.08	3.30	-0.22	2.52	7.41	15.11
1993:1	2.99	3.75	-0.76	3.16	7.15	3.32
1993:2	2.98	3.65	-0.67	2.64	7.07	12.98
1993:3	3.02	3.82	-0.80	1.51	6.80	12.56
1993:4	3.08	3.93	-0.85	1.82	6.62	7.32
1994:1	3.25	3.81	-0.56	0.99	6.56	7.10
1994:2	4.04	4.10	-0.06	2.21	6.17	-1.27
1994:3	4.51	4.28	0.23	3.49	6.00	0.93
1994:4	5.28	4.30	0.98	1.90	5.62	-0.86
1995:1	5.78	4.37	1.41	2.69	5.47	0.39
1995:2	5.62	3.98	1.65	2.46	5.68	1.17
1995:3	5.38	3.71	1.67	1.93	5.67	-3.19
1995:4	5.27	3.63	1.64	1.13	5.58	-0.56
1996:1	4.95	3.69	1.26	2.25	5.54	-2.56
1996:2	5.04	3.64	1.40	2.06	5.48	1.35
1996:3	5.14	3.91	1.22	1.50	5.27	6.17
1996:4	4.97	3.84	1.13	0.77	5.31	-3.21
1997:1	5.06	3.85	1.21	3.01	5.23	5.72
1997:2	5.07	4.08	0.99	0.23	4.98	4.45
1997:3	5.06	4.15	0.91	1.04	4.86	-1.30
1997:4	5.09	4.24	0.85	1.05	4.67	6.14
1998:1	5.05	4.23	0.82	-0.18	4.65	4.30
1998:2	4.98	4.43	0.55	1.03	4.42	3.12
1998:3	4.82	4.34	0.48	0.94	4.53	4.01
1998:4	4.25	4.39	-0.14	0.13	4.43	8.42
1999:1	4.41	4.61	-0.20	1.96	4.28	2.44
1999:2	4.45	4.94	-0.48	2.02	4.26	0.66
1999:3	4.65	4.92	-0.27	1.45	4.22	-2.80

Endnotes

1. Paul Volcker was chairman of the Federal Reserve between 1979:3 and 1987:2, but the period in question is only 1979:4 to 1982:3.

2. The data used for all estimates and tests in this article, including the data on the first-stage regressors, are available at <http://fairmodel.econ.yale.edu>. The results can be duplicated by downloading the data and some software from the site. The price variable used to construct the inflation variable is the price deflator for domestic sales. This variable was used in Fair (1978), and has been used ever since in equation 1. The three-month Treasury bill rate is used for the interest rate. Although in practice the Federal Reserve controls the federal funds rate, the quarterly average of the federal funds rate and the quarterly average of the three-month Treasury bill rate are so highly correlated that it makes little difference which rate is used in estimated interest rate rules using quarterly data. The money supply data are from the Flow of Funds Accounts of the Board of Governors of the Federal Reserve System.

3. See Fair (1994, Chapter 5) for the use of this test. The latest tests are available at <http://fairmodel.econ.yale.edu>.

4. See Fair (1994, Chapter 4) for a general discussion of these types of tests.

5. See Hendry, Pagan, and Sargan (1984).

6. The output gap measure used is $(YS - Y)/YS$, where Y is actual output and YS is a measure of potential output. These variables are in the U.S. model, which can be found at <http://fairmodel.econ.yale.edu>. YS is computed as potential productivity × potential employment, where both potential series are computed from peak-to-peak interpolations.

7. See, for example, Feldstein and Stock (1993), Hall and Mankiw (1993), Judd and Motley (1993), Clark (1994), Croushore and Stark (1994), Thornton (1995), Fair and Howrey (1996), Rudebusch (1999), Fair (2000), and Clarida, Galí, and Gertler (2000). Taylor (1985, p. 61, fn. 1) cites much of the literature prior to 1985.

8. In the following discussion, "variance" is used to refer to a particular measure of variability, not an actual variance. Variances of endogenous variables differ over time, and the variability measure is roughly an average of the quarterly variances across time.

References

Andrews, Donald W. K., and Ray C. Fair. 1988. "Inference in Nonlinear Econometric Models with Structural Change." Review of Economic Studies 55, no. 4: 615-40.

Christian, James W. 1968. "A Further Analysis of the Objectives of American Monetary Policy." Journal of Finance 23: 465-77.

Clarida, Richard, Jordi Galí, and Mark Gertler. 2000. "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." Quarterly Journal of Economics 115: 147-80.

Clark, Todd E. 1994. "Nominal GDP Targeting Rules: Can They Stabilize the Economy?" Federal Reserve Bank of Kansas City Economic Review 79, no. 3: 11-25.

Croushore, Dean, and Tom Stark. 1994. "Evaluating McCallum's Rule for Monetary Policy." Federal Reserve Bank of Philadelphia Research Working Paper no. 94-26, November.

Dewald, William G., and Harry G. Johnson. 1963. "An Objective Analysis of the Objectives of American Monetary Policy, 1952-61." In Deane Carson, ed., Banking and Monetary Studies, 171-89. Homewood, Ill.: Richard D. Irwin.

Fair, Ray C. 1978. "The Sensitivity of Fiscal Policy Effects to Assumptions about the Behavior of the Federal Reserve." Econometrica 46, no. 5: 1165-79.

———. 1994. Testing Macroeconometric Models. Cambridge, Mass.: Harvard University Press.

———. 2000. "Estimated, Calibrated, and Optimal Interest Rate Rules." May.

Fair, Ray C., and E. Philip Howrey. 1996. "Evaluating Alternative Monetary Policy Rules." Journal of Monetary Economics 38, no. 2: 173-93.

Feldstein, Martin, and James H. Stock. 1993. "The Use of a Monetary Aggregate to Target Nominal GDP." NBER Working Paper no. 4304.

Hall, Robert E., and N. Gregory Mankiw. 1993. "Nominal Income Targeting." NBER Working Paper no. 4439.

Hansen, Lars Peter. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." Econometrica 50, no. 4: 1029-54.

Hendry, David F., Adrian R. Pagan, and J. Denis Sargan. 1984. "Dynamic Specifications." In Z. Griliches and M. D. Intriligator, eds., Handbook of Econometrics, 1023-1100. Amsterdam: North-Holland.

Judd, John P., and Brian Motley. 1993. "Using a Nominal GDP Rule to Guide Discretionary Monetary Policy." Federal Reserve Bank of San Francisco Economic Review, no. 3: 3-11.

Judd, John P., and Glenn D. Rudebusch. 1998. "Taylor's Rule and the Fed: 1970-1997." Federal Reserve Bank of San Francisco Economic Review, no. 3: 3-16.

Khoury, Salwa S. 1990. "The Federal Reserve Reaction Function: A Specification Search." In Thomas Mayer, ed., The Political Economy of American Monetary Policy, 27-49. Cambridge: Cambridge University Press.

McNees, Stephen K. 1986. "Modeling the Fed: A Forward-Looking Monetary Policy Reaction Function." New England Economic Review, November-December: 3-8.

———. 1992. "A Forward-Looking Monetary Policy Reaction Function: Continuity and Change." New England Economic Review, November-December: 3-13.

Orphanides, Athanasios. 1997. "Monetary Policy Rules Based on Real-Time Data." Board of Governors of the Federal Reserve System Finance and Economics Discussion Paper Series no. 1998/03, December.

Rudebusch, Glenn D. 1999. "Is the Fed Too Timid? Monetary Policy in an Uncertain World." March.

Taylor, John B. 1985. "What Would Nominal GDP Targeting Do to the Business Cycle?" Carnegie-Rochester Series on Public Policy 22 (spring): 61-84.

References (Continued)

———. 1999. "An Historical Analysis of Monetary Policy Rules." In John B. Taylor, ed., Monetary Policy Rules, 319-44. Chicago: University of Chicago Press.

Thornton, Saranna R. 1995. "Adaptive Monetary Policy Rules Utilizing Either MZM or M2 and the Effects of Positive Bias in Money Control Errors." Paper presented at the American Economic Association Meetings, Washington, D.C., January.

James Orr, Robert Rich, and Rae Rosen

Leading Economic Indexes for New York State and New Jersey

- New indexes of leading economic indicators, presented in this article, can help predict the future course of economic activity in New York State and New Jersey.
- The leading indexes provide a basis for constructing separate indexes that estimate the probability of a recession or expansion occurring in the states within a nine-month period.
- The historical performance of the leading indexes suggests that they effectively summarize information about future economic trends beyond that offered by other indicators. The recession and expansion indexes prove particularly reliable in forecasting cyclical changes in state economic activity.

Despite the unusual length of the current expansion, few economists are ready to repudiate the business cycle. In particular, imbalances in the U.S. economy may develop rather quickly and result in either a slowdown or actual contraction in economic activity. The pattern of recurrent transitions between periods of economic expansion and contraction is of practical interest to consumers, businesses, and the government. In advance of a likely contraction, households may want to defer spending, businesses may seek to adjust product lines, and policymakers may need time to change plans and budgets. Moreover, in the midst of a lengthy contraction, retailers would like advance notice of an upturn in order to increase inventories before demand revives.

The empirical regularities that characterize the U.S. business cycle have been the subject of considerable economic research. Arthur Burns and Wesley Mitchell are among the pioneers who provided a systematic study of these features. An outcome of their research was the identification of coincident and leading indicators of economic activity. These indicators provide the basis for the development of coincident and leading indexes—composite measures intended to gauge the current level of economic activity and predict its future course. Macroeconomists, however, have expressed concern about the methodology used to construct the indexes. Because the

James Orr is a research officer, Robert Rich a senior economist, and Rae Rosen a senior economist and public information officer at the Federal Reserve Bank of New York.

The authors thank Alan Clayton-Matthews, Theodore Crone, Ray Owens, Gabriel Perez Quiros, the New York State Department of Labor, attendees at the System Committee on Regional Analysis in Minneapolis, and two anonymous referees for their valuable comments. They are also grateful to Alan Clayton-Matthews for providing the computer program used in the estimation of the index of coincident economic indicators for New York State and New Jersey. Shawn Cole, Andrei Levchenko, and Amanda Moses provided excellent research assistance. The views expressed are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

business cycle lacks a precise definition, the approach is largely judgmental and the indexes do not have well-defined statistical properties. In addition, movements in the indexes do not lend themselves to a straightforward interpretation. Although the indexes may be easy to understand at a conceptual level, it is not entirely clear what they are actually measuring.

In a series of papers, Stock and Watson (1989, 1991, 1993) attempt to provide a formal analysis of coincident and leading indexes for the United States. Incorporating the idea that the U.S. business cycle represents broad-based contractions and expansions in economic activity, the authors develop a model that estimates a common unobserved factor across a set of coincident indicators. This common factor is assumed to represent the shared influence of "the state of the economy" on the indicators, and is identified as the coincident index. Because the leading index is designed to predict changes in "the state of the economy," Stock and Watson define the leading index as the forecasted growth in the coincident index.

Coincident and leading indexes [are] composite measures intended to gauge the current level of economic activity and predict its future course.

If national business cycles mirrored those at the state level, then the indexes discussed above might be sufficient for summarizing and forecasting regional economic activity. However, economic fluctuations at the national level are not reflected evenly throughout the fifty states. During the 1980s and 1990s, the nation's recessions and expansions were characterized as bicoastal, led by the "rust belt," and, later still, led by the revitalized "industrial heartland." Different regions of the country clearly led and lagged changes in the nation's economy. Moreover, while single variables are often used as shorthand measures for gauging the current level or future course of state economic activity, they may be too narrow in scope or released too late to serve as a useful guide.[1]

This article describes a method by which we may more accurately predict regional economic activity. Specifically, we develop an index of leading economic indicators (LEI) for New York State and for New Jersey over the 1972-99 period. We extend our earlier work (Orr, Rich, and Rosen 1999), which uses the dynamic factor model of Stock and Watson (1989, 1991) to estimate an index of coincident economic indicators (CEI) for each state and to date each state's business cycles. We define the LEI as the forecasted growth in each state's CEI over a nine-month horizon. The forecasts are constructed using a single-equation model, where the set of leading indicators includes the national index of leading economic indicators and an interest rate spread as well as state-level changes in the CEI and housing permits.

We then develop a recession index and an expansion index for New York State and for New Jersey. The indexes estimate the probability of a recession or expansion in each state over the next nine months. To construct the indexes, we define recessionary and expansionary patterns in terms of future growth sequences in the CEI. We then calculate the indexes by using simulation techniques—Monte Carlo methods—to evaluate the likelihood of observing the recessionary and expansionary patterns.

We find that the movements of our recession and expansion indexes show a close relationship with the behavior of our LEI for New York State and New Jersey. The indexes therefore allow us to extend the informational content of our LEI by estimating the probability of an upcoming cyclical change in each state's economic activity.

In the next section, we provide a brief history and description of state indexes of leading economic indicators. We then discuss the construction of our leading indexes for New York State and New Jersey and provide details on the methodology used to estimate our recession and expansion indexes. Finally, we present the empirical results and examine the within-sample and out-of-sample performance of the indexes.

Existing State Indexes of Leading Economic Indicators

Although a national index of leading economic indicators has been developed to signal future turning points in aggregate economic activity, similar indexes at the state level are not widely available. In addition to the indexes developed in this article, leading indexes are regularly reported for only three states—New Jersey, Pennsylvania, and Texas.[2]

A key factor constraining the evaluation of leading indexes at the state level is that, unlike the nation, states do not have formally designated business cycles.[3] Therefore, an index of coincident economic indicators for each state is usually constructed prior to the development of an index of leading economic indicators.[4] Peaks and troughs in the coincident index can then be used to date state business cycles.[5] Monthly coincident and leading indexes for Texas were developed in 1988 and refined in 1990 (Phillips 1988, 1990). Since 1994, the Federal Reserve Bank of Philadelphia has published monthly indexes of coincident economic indicators for Pennsylvania and New Jersey and introduced an index of leading economic indicators for the two states in 1996 (Crone 1994; Crone and Babyak 1996).[6]

Important issues in the construction of a leading index include the selection of variables and the method of combining the variables into a single composite measure. For Texas, the index of leading economic indicators is constructed as a weighted average of leading variables. A variable is included in the leading index if changes in its past values (of at least two months) are highly correlated with current changes in the coincident index. The resulting leading index contains a total of nine variables, six of which are: two state counterparts to variables entering the national index of leading economic indicators (average weekly hours worked in manufacturing and initial claims for unemployment insurance); two variables that point to the future performance of the oil industry (new well permits and real oil prices); an indicator of international competitiveness (a Texas trade-weighted value of the dollar); and the national index of leading economic indicators.[7] Three other state variables are included: a state help-wanted index, real retail sales, and an index of the real stock prices of Texas-based companies. Each variable is assigned a weight that emphasizes the cyclical timing of the series. This procedure is similar to the one used by the U.S. Department of Commerce in assigning weights to the variables entering the national index of leading economic indicators.

Although a national index of leading economic indicators has been developed to signal future turning points in aggregate economic activity, similar indexes at the state level are not widely available.

For Pennsylvania and New Jersey, Crone and Babyak (1996) adopt a forecasting approach, similar to the one used by Stock and Watson (1989), to construct an index of leading economic indicators. The leading index for each state is a forecast of the change in its coincident index over the next nine months. For Pennsylvania, the forecasting equation includes four state variables: the number of new housing permits, initial claims for unemployment insurance, an index of local vendor delivery time, and lagged values of the coincident index; and an interest rate spread measuring the difference between the rates on ten-year Treasury bonds and one-year Treasury bills. Similar variables are used in the construction of the New Jersey LEI, except that the vendor delivery index is excluded and the interest rate spread measures the difference between the rates on six-month commercial paper and six-month Treasury bills. The weight of each variable in the leading index is determined by its estimated coefficient in the forecasting equation.

For the purpose of forecasting recessions and expansions, a procedure must be established that translates movements in the leading index into a signal about future turning points in economic activity. Crone and Babyak (1996) adopt a rule of thumb in which three consecutive negative (positive) readings of the leading index signal an upcoming recession (expansion).[8] Based on that rule, the indexes for Pennsylvania and New Jersey have done reasonably well in forecasting recessions, although not as well in forecasting expansions.[9]

Phillips (1990) uses an alternative approach that converts the current reading of the Texas LEI into the probability of a recession or expansion. The probability is computed by calculating the likelihood that the current reading of the LEI would occur during a recession or expansion and then modifying the value based on the probability for the previous period. A probability greater than 90 percent is taken as a strong signal of an impending recession or expansion.[10]

Indexes of Future Economic Activity for New York State and New Jersey

This section describes the construction of the indexes of leading economic indicators for New York State and New Jersey as well as the recession and expansion indexes. We begin by specifying a model that links the LEI to near-term growth in the coincident index for each state. The analysis then turns to a discussion of the use of the LEI to forecast recessions and expansions. To motivate the design of the recession and expansion indexes, we initially examine a popular rule of thumb that uses the behavior of the LEI to signal future recessions and expansions. To improve upon its features and performance, we propose a slight modification to this rule. We argue that one advantage of our modification is an increased lead time in generating a recession or expansion signal. More importantly, however, our modification provides the basis for the construction of the recession index and expansion index, which estimate, respectively, the probability of a future recession or expansion. Unlike conventional rules of thumb, the indexes are defined over a continuous probability range.

It is worth noting that we rely on different definitions of recessions and expansions for two aspects of the analysis. In the absence of a timely and fully reliable indicator of future recessions and expansions at the state level, we generate forecasts of these events by defining them on an *ex ante* basis in terms of future growth sequences in the coincident index. However, we date state business cycles, and hence evaluate the performance of the recession/expansion forecasts, by identifying peaks and troughs in the coincident index on an *ex post* basis. Although our approach might suggest some inconsistency, it parallels the approach often taken at the national level. For example, the NBER's dating of recessions normally occurs some time after a turning point has been realized.[11] In addition, the lack of a timely and fully reliable indicator of future recessions and expansions at the national level typically requires the adoption of some definition of recessions and expansions for forecasting purposes.[12]

The Leading Economic Index

Following the analyses of Stock and Watson (1989) and Crone and Babyak (1996), the indexes of leading economic indicators for New York State and New Jersey are forecasts of the change in the state's index of coincident economic indicators. Chart 1 extends the work of Orr, Rich, and Rosen (1999) and plots the CEI for each state, where the indexes are constructed using the methodology developed in Stock and Watson (1989, 1991) and the shading indicates state recessions.[13] As discussed in Orr, Rich, and Rosen, the cycles for New York State and New Jersey are broadly similar to those at the national level, although there are marked differences across some episodes with regard to timing, duration, and magnitude.

To construct the LEI, we select the nine-month forecasting horizon adopted in Crone and Babyak (1996) and specify the following regression equation:

$$(1) \qquad \Delta C_t^{t+9} = X_t \beta + \varepsilon_{t,9} \ ,$$

where ΔC_t^{t+9} is the (annualized) growth in the CEI between month t and month t+9, X_t is a set of leading variables available in month t used to forecast ΔC_t^{t+9}, β denotes the coefficients relating the leading variables to future growth in the CEI, and $\varepsilon_{t,9}$ is a mean-zero error term.[14]

By its definition, the LEI can be recovered from the estimation of equation 1 as:

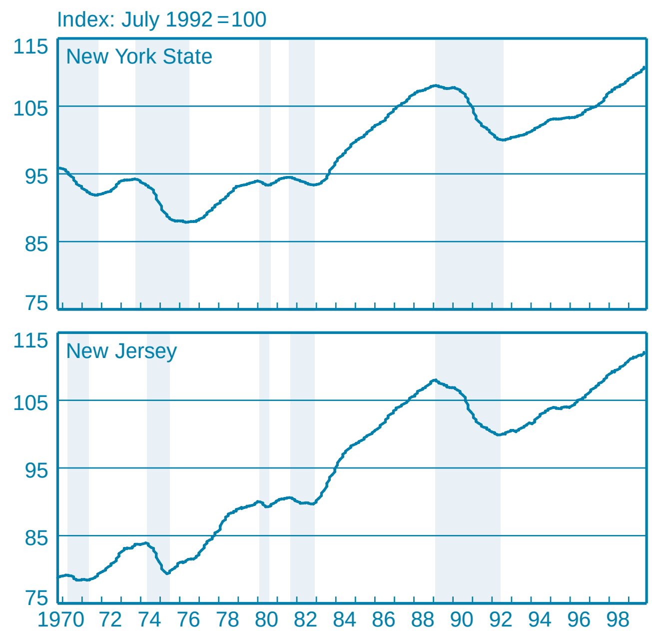
$$(2) \qquad \Delta \hat{C}_t^{t+9} = X_t \hat{\beta},$$

where $\Delta \hat{C}_t^{t+9}$ is the predicted nine-month growth rate in the CEI in month t and $\hat{\beta}$ denotes the estimated coefficients of the model.

Before turning our attention to the issue of forecasting recessions and expansions, there are two points about the leading index and our methodology that merit discussion. First, the single-equation approach contrasts with Stock and Watson (1989) and Crone and Babyak (1996), who use a vector autoregression (VAR) model to construct the leading index.[15] An advantage of our framework is that it is tractable and very easy to interpret. Specifically, it allows us to consider a direct relationship between the quantity of interest ($\Delta \hat{C}_t^{t+9}$) and observed values of the leading variables, rather than relying on a sequence of forecasted values that must be derived in a recursive manner. In addition, equation 1 involves the estimation of a relatively small number of parameters compared with a VAR system. This is consistent with the principle of parsimony and helps reduce the risk of overfitting the data.

Second, there is a technical issue that arises from the use of overlapping data in equation 1. Specifically, the forecasting horizon of the CEI exceeds the sampling interval of the data. Consequently, the error term in equation 1 is not precluded from displaying autocorrelation. Although this feature of the data does not invalidate the use of conventional techniques to estimate the parameters of equation 1, the standard errors need to be calculated using methods designed to account for autocorrelation of the disturbance terms.[16]

Chart 1
Indexes of Coincident Economic Indicators
October 1969–November 1999



Source: Authors' calculations.

Note: The shading indicates state recessions as determined by the authors.

The Recession and Expansion Indexes

The central reason for developing a leading index is to obtain a series capable of providing a reliable signal about recessions and expansions. It is important to note that by itself, a leading index conveys only qualitative information about the future course of fluctuations in economic activity. That is, its design is primarily intended to generate turning points that precede those associated with the business cycle. From a quantitative perspective, however, a leading index does not offer a probabilistic forecast of recessions and expansions.

Researchers have attempted to provide a more formal link between leading indexes and the incidence of recessions and expansions in the economy. While a review of the literature is beyond the scope of this article, it is instructive to examine what might be considered the two most divergent approaches.

We derive measures indicating the likelihood of a recession or expansion in a state—the recession and expansion indexes—by using simulation techniques to estimate the probability of observing the [recessionary and expansionary] growth patterns.

One approach is to specify a rule of thumb that uses movements in a leading index to forecast recessions and expansions. Crone and Babyak (1996) adopt this procedure in their work on leading indexes for Pennsylvania and New Jersey. The authors define the leading index as the forecasted nine-month growth rate in the coincident index, and apply their rule of thumb to the behavior of the leading index over three consecutive months. In particular, the rule generates a recession (expansion) signal if the economy is currently in an expansion (recession) and there are three consecutive declines (increases) in the leading index. An advantage of this approach is that it is very easy to implement. A potential drawback is that the rule is arbitrary and implicitly associates forecasts of recessions and expansions with probabilities that are restricted to being either 0 or 1. That is, the rule does not allow a "partial" recession (or expansion) signal to be issued.

The approach of Stock and Watson (1989, 1993) represents the other extreme. Stock and Watson define recessions and expansions in terms of particular sequences of one-month growth rates in the coincident index and then generate forecasts of recessions and expansions by evaluating the likelihood of observing the sequences over a six-month horizon. An attractive feature of the framework is that it offers a statistical model to construct probabilistic statements about future recessions and expansions. Its main disadvantage is that the model is quite sophisticated and not easily amenable to predicting recessions and expansions over alternative horizons.[17]

Our approach to forecasting recessions and expansions for New York State and New Jersey essentially combines the approaches of Stock and Watson (1989, 1993) and Crone and Babyak (1996). Specifically, we draw upon the work of Stock and Watson by defining recessions and expansions in terms of future growth patterns in the CEI. In addition, we derive measures indicating the likelihood of a recession or expansion in a state—the recession and expansion indexes—by using simulation techniques to estimate the probability of observing the growth patterns. In contrast to Stock and Watson, however, we employ simpler definitions for the recessionary and expansionary patterns based on a rule-of-thumb formulation. Because of these considerations, our approach can be interpreted as a "probabilistic rule of thumb."[18]

To motivate our modeling strategy, we begin by examining the rule of thumb used by Crone and Babyak to signal a recession or expansion:

$$(3)\ (\Delta\hat{C}_t^{t+9}<0 \textit{ and } \Delta\hat{C}_{t+1}^{t+10}<0 \textit{ and } \Delta\hat{C}_{t+2}^{t+11}<0)\ \textit{[recession signal]}$$

$$(\Delta\hat{C}_t^{t+9}>0 \textit{ and } \Delta\hat{C}_{t+1}^{t+10}>0 \textit{ and } \Delta\hat{C}_{t+2}^{t+11}>0)\ \textit{[expansion signal]},$$

where $\Delta\hat{C}_{t+i}^{t+i+9}$ denotes the forecasted growth rate in the CEI between month $t+i$ and month $t+i+9$ using information available through month $t+i$.

As an alternative to equation 3 and its reliance on three consecutive readings of a (nine-month) leading index, we will initially consider a rule that uses readings of forecasted growth in the CEI over three adjacent horizons to signal a recession or expansion:

$$(4)\ (\Delta\hat{C}_t^{t+7}<0 \textit{ and } \Delta\hat{C}_t^{t+8}<0 \textit{ and } \Delta\hat{C}_t^{t+9}<0)\ \textit{[recession signal]}$$

$$(\Delta\hat{C}_t^{t+7}>0 \textit{ and } \Delta\hat{C}_t^{t+8}>0 \textit{ and } \Delta\hat{C}_t^{t+9}>0)\ \textit{[expansion signal]},$$

where $\Delta\hat{C}_t^{t+i}$ denotes the forecasted growth rate in the CEI between month t and month $t+i$ using information available through month t. Because the leading index is linked to predicted growth in the CEI, equation 4 can be interpreted as a rule of thumb that uses concurrent declines (increases) in a seven-, eight-, and nine-month LEI to predict future recessions (expansions).

The key differences between equations 3 and 4 concern the dating of the information sets used for forecasting and the immediacy with which a recession or expansion signal can be generated. If $\Delta\hat{C}_t^{t+9}<0$ ($\Delta\hat{C}_t^{t+9}>0$), then the rule in equation 3 not only requires two additional periods to pass before a signal can be issued in period t+2, but also stipulates that the signal partly relies on forecasts constructed during the previous two periods. In contrast, the rule in equation 4 allows an immediate signal to be issued in period t because it depends only on currently available information. An advantage of the formulation in equation 4 is that it can improve the lead time in signaling a transition from one phase of the business cycle to the other, although it might be more susceptible to generating false signals.

If our interest was restricted to predicting recessions and expansions based on the rules of thumb in equations 3 and 4, then the remainder of the analysis might be concerned simply with evaluating the rules' relative performance. However, both equations are unattractive because they imply a discrete probability pattern for recessions and expansions that admits only values of 0 or 1. Moreover, the signal (and probability) of a recession or expansion depends only on the *sign* of the forecasted growth rates in the CEI and not on the *magnitude*.

In an attempt to remedy both of these shortcomings, we develop a formal statistical model to forecast recessions and expansions. Following Stock and Watson (1989, 1993), we define recessionary and expansionary patterns in terms of a sequence of growth rates in the CEI. In particular, we extend equation 4 and adopt the following definitions:

(5) $(\Delta C_t^{t+7} < 0 \text{ and } \Delta C_t^{t+8} < 0 \text{ and } \Delta C_t^{t+9} < 0)$ *[recessionary pattern]*

$(\Delta C_t^{t+7} > 0 \text{ and } \Delta C_t^{t+8} > 0 \text{ and } \Delta C_t^{t+9} > 0)$ *[expansionary pattern]*,

where ΔC_t^{t+i} denotes the actual rate of growth in the CEI between month t and month $t+i$.[19]

For the purpose of forecasting recessions and expansions, equation 5 is void of any operational content because it is expressed in terms of *future* growth rates in the CEI. However, if we view the sequences in the equation as realizations from a stochastic process, then we can use the associated probabilities as the basis for drawing inferences about the likelihood of a recession or expansion. Borrowing from the terminology of Stock and Watson (1989, 1993), we define the recession index (R_t) and expansion index (E_t) in month t as:

(6) $$R_t = Pr\,[(\Delta C_t^{t+7} < 0 \text{ and } \Delta C_t^{t+8} < 0 \text{ and } \Delta C_t^{t+9} < 0)|I_t]$$

$$E_t = Pr\,[(\Delta C_t^{t+7} > 0 \text{ and } \Delta C_t^{t+8} > 0 \text{ and } \Delta C_t^{t+9} > 0)|I_t],$$

where R_t and E_t denote, respectively, the probability of a recession and expansion within the next nine months, conditional on available information through month t (I_t).

We complete the statistical model by augmenting the previous specification for the nine-month growth rate in the CEI in equation 1 to include the processes governing the movements in the CEI over a seven- and eight-month horizon:

(7) $$\Delta C_t^{t+7} = X_{1t}\beta_1 + \varepsilon_{t,7}$$

$$\Delta C_t^{t+8} = X_{2t}\beta_2 + \varepsilon_{t,8}$$

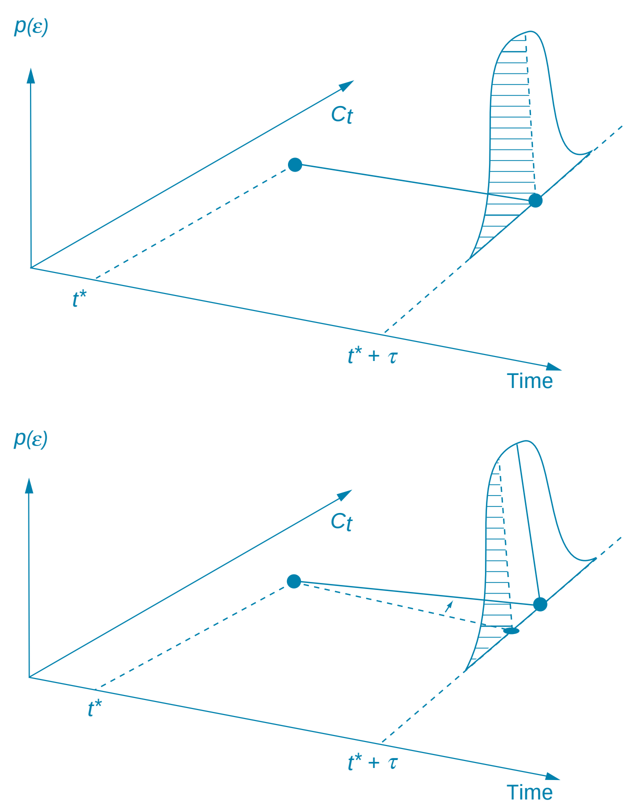
$$\Delta C_t^{t+9} = X_{3t}\beta_3 + \varepsilon_{t,9},$$

where X_{it} is a set of leading variables in the ith equation available in month t, β_i refers to the coefficients in the ith equation, and $\varepsilon_{t,j}$ is a mean-zero error term associated with the j-month-ahead growth rate in C_t. Our technical appendix describes the specification of the system in equation 7 for New York State and New Jersey and provides details on the calculation of the recession and expansion indexes in equation 6.[20]

With regard to forecasting recessions and expansions, the key aspect of our approach is that it depends on the processes governing the deterministic growth component of the CEI *and* the random disturbance term.[21] As such, our statistical model takes into account forecasting uncertainty and incorporates gradations of forecasted growth rates into the formulation of the recession and expansion indexes.

The top panel of Chart 2 makes this point visually by depicting a hypothetical situation in which the CEI is expected to remain constant (zero growth rate) over the near term. For

Chart 2
Calculating the Recession Index



the purpose of illustration, we will assume that a rule of thumb and our recession (and expansion) index depend only on the behavior of the CEI over a singular horizon.[22]

A rule of thumb based solely on the sign of the forecasted growth in the CEI between period t^* and period $t^* + \tau$ would not generate a recession signal and would implicitly assign a value of zero to the likelihood of a future recession. In contrast, our approach would recognize that there is an equal likelihood of positive and negative growth rates for the CEI and would assign a probability of 50 percent (indicated by the shaded area) to a future recession. The bottom panel of the chart then illustrates how the recession signal from a rule of thumb would be invariant to higher (positive) predicted growth in the CEI, while the value of our recession index would decline because of the lower probability of observing negative growth in the CEI.[23]

Our statistical model takes into account forecasting uncertainty and incorporates gradations of forecasted growth rates into the formulation of the recession and expansion indexes.

Empirical Results

The construction of the indexes for New York State and New Jersey requires the selection of a set of leading indicators. Our list includes both state-level and national variables. The state-level data consist of past changes in the coincident index and housing permits.[24] The use of lagged growth rates of the coincident index is intended to capture inertial effects—persistence—in the series. We augment these explanatory variables by including past changes in the national index of leading economic indicators and an interest rate spread—the difference between the yields on ten-year Treasury bonds and one-year Treasury bills. Our data appendix provides further details on the leading indicators.

It is worth noting that the results indicate that the interest rate spread contains additional forecasting power for state economic activity despite the inclusion of financial market indicators in the national leading index.[25] This evidence documenting an independent role for the interest rate spread as a regional indicator may reflect the critical importance of the finance, insurance, and real estate (FIRE) industry to the economies of New York State and New Jersey. For example, the FIRE industry's shares of both employment and earnings in the region are much larger than they are in the national economy. In addition, Kuttner and Sbordone (1997) provide evidence that shocks to employment growth in the FIRE sectors of New York State and New Jersey have significant effects on overall employment growth in the region.

Within-Sample Performance of the Leading Index

In estimating the leading indexes for New York State and New Jersey, it is important to recognize the existence of publication lags in variables and the implication for model specification. We follow the conventional practice of assuming a one-month delay in the release of the coincident index for each state.[26] As a consequence, there is also a one-month delay in the release of the leading index. That is, the LEI released at the end of February would correspond to the forecasted growth in the CEI from January to September, conditional on information available through February. To help clarify the notation, we define the lags of each leading indicator relative to the current information set. For the LEI released at the end of February, the contemporaneous value (zero lag) of the coincident index, housing permits, and the national index of leading economic indicators would reflect observations through January, while the contemporaneous value of the interest rate spread would be for February. The use of the most currently available data results in a more timely measure of economic activity.

Because of the large number of possible specifications for equation 1, we applied various testing procedures to help with model selection. The specifications were evaluated based on their within-sample as well as out-of-sample performance.[27] We also examined the specifications to determine if the estimated coefficients generally displayed the "correct" sign. While the rankings based on the various criteria were not always in complete agreement, they were broadly consistent.

Our results from estimating equation 1 for New York State and New Jersey over the sample period February 1972 to November 1999 appear, respectively, in Tables 1 and 2. The leading indexes for each state appear in Charts 3 and 4, where the series are plotted based on the dating of the LEI series and the shading represents state recessions.[28] As shown by the values of the adjusted R^2 measure, the models fit the data quite well and are able to explain approximately 75 percent and 63 percent of the total variation in the nine-month CEI growth rates for New York State and New Jersey, respectively.

Table 1

Estimated Leading Economic Index Model for New York State

$$\Delta C_t^{t+9} = \alpha_0 + \sum_{i=0}^{n_1} \beta_i \Delta C_{t-i-9}^{t-i} + \sum_{i=0}^{n_2} \delta_i \Delta Permits_{t-i} + \psi_0 (Spread_t) + \sum_{i=0}^{n_3} \gamma_i \Delta LEI_{t-i}^{National} + \varepsilon_{t,9},$$

where:

$\Delta C_t^{t+9} = (1200/9)*[1n(C_{t+9}/C_t)]=$ nine-month growth in the index of coincident economic indicators (CEI),

$\Delta C_{t-i-9}^{t-i} = (1200/9)*[1n(C_{t-i}/C_{t-i-9})]$,

$\Delta Permits_t = (1200/9)*[1n(Permits_t/Permits_{t-9})]=$ nine-month growth in housing permits,

$Spread_t = (1/3)\sum_{i=0}^{2}[Yield_{t-i}^{Ten-year\ Treasury\ bond} - Yield_{t-i}^{One-year\ Treasury\ bill}]$,

$\Delta LEI_t^{National} = (1200)*[1n(LEI_t^{National}/LEI_{t-1}^{National})]=$ one-month growth in the national index of leading economic indicators (LEI).

Sample Period: February 1972-November 1999 $\bar{R}^2 = 0.749$ DOF = 326

Parameter	Coefficient	Standard Error	*t*-Statistic
α_0	-0.1318	0.1686	0.7817
β_0	1.5787	0.3653	4.3216***
β_1	-0.9379	0.3563	2.6326***
δ_0	0.0022	0.0016	1.3163
δ_3	0.0025	0.0017	1.4649
δ_6	0.0039	0.0014	2.7966***
ψ_0	0.2618	0.0688	3.8034***
γ_0	0.2425	0.0412	5.8904***

New York State	Recession Lead(+)/Lag(-)			Expansion Lead(+)/Lag(-)	
Business Cycle Peak	Conventional Rule of Thumb	Modified Rule of Thumb	Business Cycle Trough	Conventional Rule of Thumb	Modified Rule of Thumb
October 1973	+2 months	+4 months	June 1976	+4 months	+6 months
February 1980	+5 months	+7 months	August 1980	0 months	+2 months
August 1981	+0 months	+2 months	November 1982	0 months	+2 months
February 1989	-3 months	-1 month	July 1992	-4 months	-2 months
False signals	August-September 1992	August-September 1992 January 1996	False signals	March 1989	March 1989

Source: Authors' calculations.

Notes: In the top panel, all growth rates are calculated on an annual percentage basis. Standard errors are calculated using the Newey-West (1987) estimator and allow for a moving-average (8) process for the regression residuals. DOF is degrees of freedom.

In the bottom panel, the identification of peaks and troughs in the state business cycles is based upon the coincident index derived using a full-sample smoother ($C_{t|T}$). State recessions (expansions) are dated from one month after a peak (trough) to the trough (peak) of the state's coincident economic activity index in any business cycle. For the conventional rule of thumb, three consecutive declines (increases) in the nine-month index forecast are used to signal a recession (expansion) within the next nine months. For the modified rule of thumb, a simultaneous decline (increase) in the seven-, eight-, and nine-month index forecast is used to signal a recession (expansion) within the next nine months. A lead/lag value of zero months indicates that the signal coincided with the beginning of a recession/expansion.

*** Significant at the 1 percent level.

Table 2

Estimated Leading Economic Index Model for New Jersey

$$\Delta C_t^{t+9} = \alpha_0 + \sum_{i=0}^{n_1} \beta_i \Delta C_{t-i-9}^{t-i} + \sum_{i=0}^{n_2} \delta_i \Delta Permits_{t-i} + \psi_0(Spread_t) + \sum_{i=0}^{n_3} \gamma_i LEI_{t-i}^{National} + \varepsilon_{t,9},$$

where:

$\Delta C_t^{t+9} =$ (1200/9)*[1n(C_{t+9}/C_t)]=nine-month growth in the index of coincident economic indicators (CEI),

C_{t-i-9}^{t-i} (1200/9)*[1n(C_{t-i}/C_{t-i-9})],

$\Delta Permits_t =$ (1200/9)*[1n$(Permits_t/Permits_{t-9})$]=nine-month growth in housing permits,

$Spread_t = (1/3)\sum_{i=0}^{2}[Yield_{t-i}^{Ten\text{-}year\ Treasury\ bond} - Yield_{t-i}^{One\text{-}year\ Treasury\ bill}]$,

$\Delta LEI_t^{National} =$ (1200)*[1n$(LEI_t^{National}/LEI_{t-1}^{National})$]=one-month growth in the national index of leading economic indicators (LEI).

Sample Period: February 1972-November 1999 $\bar{R}^2 = 0.626$ DOF = 325

Parameter	Coefficient	Standard Error	*t*-Statistic
α_0	0.2693	0.3052	0.8823
β_0	0.5598	0.2917	1.9191*
β_1	-0.8696	0.3337	2.6063***
β_2	0.7066	0.2542	2.7795***
δ_0	0.0078	0.0032	2.4665**
δ_3	0.0041	0.0018	2.2355**
δ_6	0.0029	0.0018	1.6320
ψ_0	0.3436	0.1163	2.9536***
γ_0	0.2696	0.0723	3.7269***

New Jersey	Recession Lead(+)/Lag(-)			Expansion Lead(+)/Lag(-)	
Business Cycle Peak	Conventional Rule of Thumb	Modified Rule of Thumb	Business Cycle Trough	Conventional Rule of Thumb	Modified Rule of Thumb
May 1974	+4 months	+6 months	June 1975	0 months	+1 month
February 1980	+5 months	+7 months	July 1980	-1 month	+1 month
September 1981	-1 month	+1 month	November 1982	+4 months	+6 months
February 1989	-4 months	0 months	May 1992	+6 months	+7 months
False signals	—	—	False signals	January 1990	September 1989 December 1989 January 1990 March 1990

Source: Authors' calculations.

Notes: In the top panel, all growth rates are calculated on an annual percentage basis. Standard errors are calculated using the Newey-West (1987) estimator and allow for a moving-average (8) process for the regression residuals. DOF is degrees of freedom.

In the bottom panel, the identification of peaks and troughs in the state business cycles is based upon the coincident index derived using a full-sample smoother ($C_{t|T}$). State recessions (expansions) are dated from one month after a peak (trough) to the trough (peak) of the state's coincident economic activity index in any business cycle. For the conventional rule of thumb, three consecutive declines (increases) in the nine-month index forecast are used to signal a recession (expansion) within the next nine months. For the modified rule of thumb, a simultaneous decline (increase) in the seven-, eight-, and nine-month index forecast is used to signal a recession (expansion) within the next nine months. A lead/lag value of zero months indicates that the signal coincided with the beginning of a recession/expansion.

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

The results also confirm the presence of inertia in the growth pattern of the CEI, as evident by the quantitative and statistical significance of its past changes.

It is also important to note that the other leading variables generally are significant and always display the anticipated sign. In particular, a decline in the growth of housing permits, a narrowing of the yield spread, and lower growth in the national index of leading economic indicators are taken as harbingers of a slowing in state economic activity. Because the month-to-month movements in the national index of leading economic indicators can be somewhat noisy, we elected to smooth the series using a filter suggested by Stock and Watson (1989) to improve the model's forecasting performance.[29]

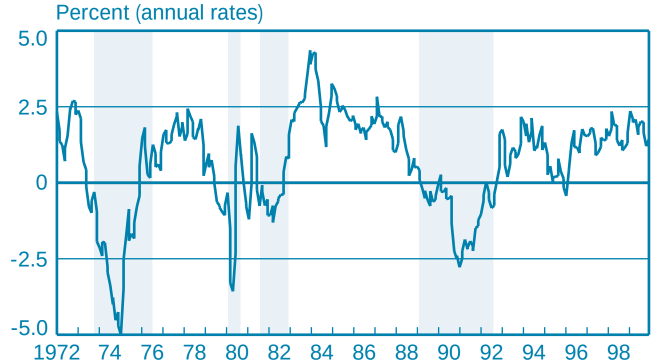
Although the New York State LEI and New Jersey LEI are expressed in terms of (annualized) growth rates rather than levels, we can nevertheless examine Charts 3 and 4 and conduct an informal and preliminary evaluation of the indexes. Specifically, if the index is providing a timely and useful signal of transitional shifts in business cycle phases, then we should observe negative (positive) growth preceding the onset of a recession (expansion).

The indexes for New York State and New Jersey generally display this feature, although there is considerable variation in the timing both within and across the states. Our inspection reveals that the indexes turned negative well in advance of the impending recessions in the mid-1970s and in 1980. Interestingly, both indexes are characterized by a fluctuating pattern prior to the 1981 recession in which they first turn negative, then increase above zero before turning negative again. In addition, each index showed an upturn prior to the start of the expansions in the mid-1970s and during the 1980s.

Chart 3

New York State Index of Leading Economic Indicators

Nine-Month CEI Forecast: January 1972–November 1999



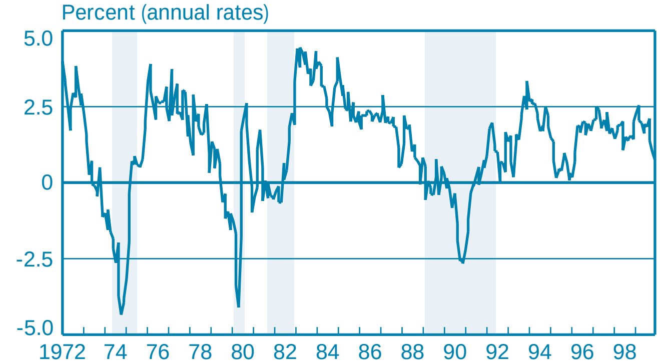
Source: Authors' calculations.

Notes: CEI is index of coincident economic indicators. The shading indicates state recessions as determined by the authors.

Chart 4

New Jersey Index of Leading Economic Indicators

Nine-Month CEI Forecast: January 1972–November 1999



Source: Authors' calculations.

Notes: CEI is index of coincident economic indicators. The shading indicates state recessions as determined by the authors.

However, the downturn in the LEI for both states was subsequent to the onset of the recessions in 1989.[30] With regard to their recent behavior, the series have displayed a similar pattern and were characterized by very slow growth during the mid-1990s.

To gain further insight into the behavior of the LEI, we can consider an historical decomposition to isolate each indicator's contribution to the index. The historical decomposition is based on the following relationship:

$$(8) \qquad \Delta \hat{C}_t^{t+9} = \Delta \bar{C}_t^{t+9} + (X_t - \bar{X})\hat{\beta},$$

which expresses the LEI as the mean nine-month growth rate in the CEI ($\Delta \bar{C}_t^{t+9}$) plus the sum of the contributions from each indicator. Charts 5 and 6 plot the (mean-adjusted) LEI and its historical decomposition for New York State and New Jersey. For convenience, we accumulate the effects from all lagged terms when calculating the total contribution of a series.[31] A positive (negative) value for a series indicates that the variable is contributing to greater than (less than) trend growth in the leading index.

As shown, each of the series makes a contribution to the total. The largest historical contributions are from past changes in the coincident index and the national index of leading economic indicators. While the contributions from the yield spread and housing permits are smaller in scope, both variables have the desirable property of displaying downturns/upturns that precede those associated with state business cycles. Also, it appears that housing permits contribute more to the LEI for New Jersey than for New York State.

Before turning our attention to the recession and expansion indexes, it is instructive to examine the properties of the recessionary and expansionary patterns that underlie their construction. To shed light on this issue, we explore the consequences of forecasting recessions and expansions based on the rules of thumb discussed earlier. The conventional rule of thumb described in equation 3 is given by:

$$(\Delta\hat{C}_t^{t+9} < 0 \text{ and } \Delta\hat{C}_{t+1}^{t+10} < 0 \text{ and } \Delta\hat{C}_{t+2}^{t+11} < 0) \text{ [recession signal]}$$

$$(\Delta\hat{C}_t^{t+9} > 0 \text{ and } \Delta\hat{C}_{t+1}^{t+10} > 0 \text{ and } \Delta\hat{C}_{t+2}^{t+11} > 0) \text{ [expansion signal]}.$$

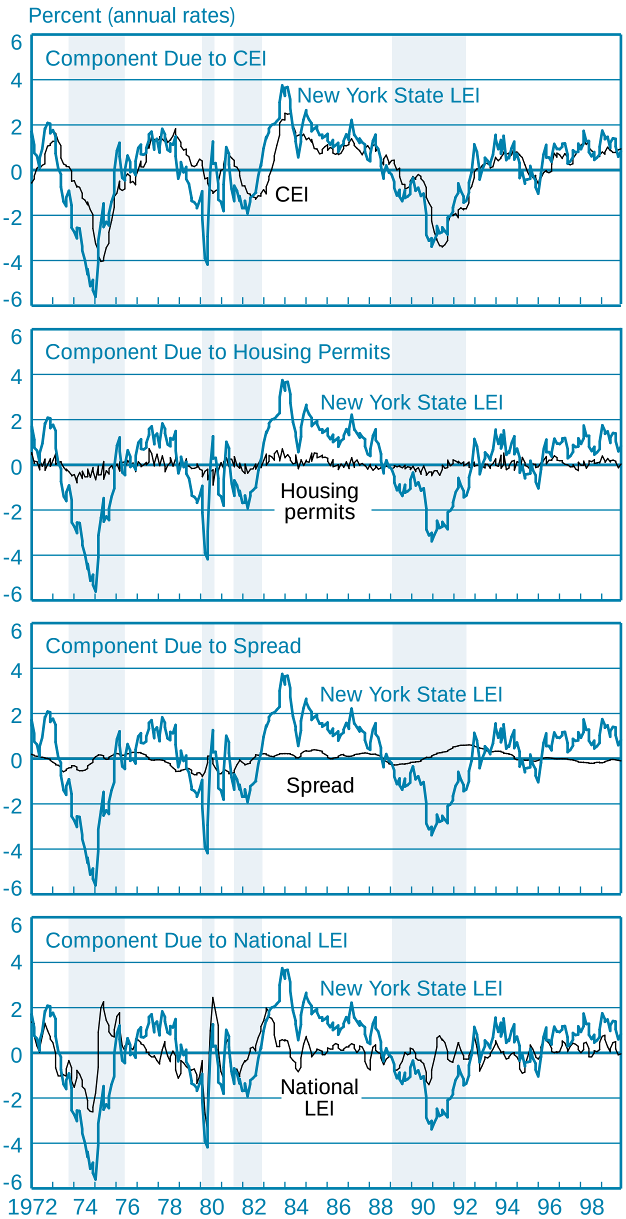
As an alternative, we propose the following modified rule of thumb in equation 4:

$$(\Delta\hat{C}_t^{t+7} < 0 \text{ and } \Delta\hat{C}_t^{t+8} < 0 \text{ and } \Delta\hat{C}_t^{t+9} < 0) \text{ [recession signal]}$$

$$(\Delta\hat{C}_t^{t+7} > 0 \text{ and } \Delta\hat{C}_t^{t+8} > 0 \text{ and } \Delta\hat{C}_t^{t+9} > 0) \text{ [expansion signal]}.$$

Chart 5

Historical Decomposition of the New York State LEI

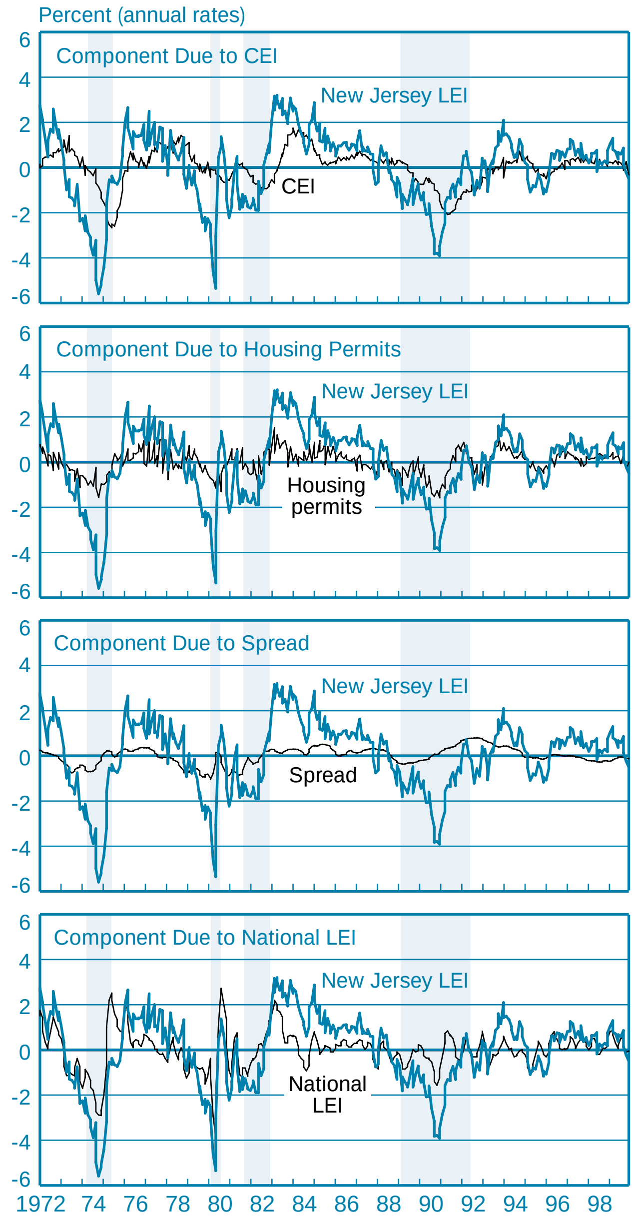


Source: Authors' calculations.

Notes: LEI is index of leading economic indicators; CEI is index of coincident economic indicators. The shading indicates state recessions as determined by the authors.

Chart 6

Historical Decomposition of the New Jersey LEI



Source: Authors' calculations.

Notes: LEI is index of leading economic indicators; CEI is index of coincident economic indicators. The shading indicates state recessions as determined by the authors.

Tables 1 and 2 also compare the forecasting performance of these competing rules of thumb. An initial examination of the results indicates that both rules ultimately signaled every recession and expansion over the sample period. In those instances where they failed to provide a leading signal, the rules nevertheless generated an accurate, albeit lagging, signal of an expansion or recession. The traditional rule of thumb was only able to provide an accurate leading signal for half of the recessions and expansions across New York State and New Jersey. With regard to the modified rule, it provided an accurate leading signal of three of the four recessions in New York State as well as three of the four expansions.[32] For New Jersey, the modified rule provided an accurate leading signal of three of the four recessions and all four expansions. The results for predicting expansions based on the modified rule seem particularly noteworthy in light of the previous lack of success at the state level.

The findings also seem to confirm our earlier suspicion about the properties of the two rules. The modified rule generates an earlier signal of recessions and expansions, where the lead time of equation 4 typically exceeds that of equation 3 by two months. These results are not particularly surprising given differences in the formulation of the rules and in the dating of information sets. These same considerations would also seem to account for the modified rule being slightly more susceptible to issuing false signals, although this feature seems to be principally limited to the case of forecasting the recovery in New Jersey in early 1990.

Within-Sample Performance of the Recession and Expansion Indexes

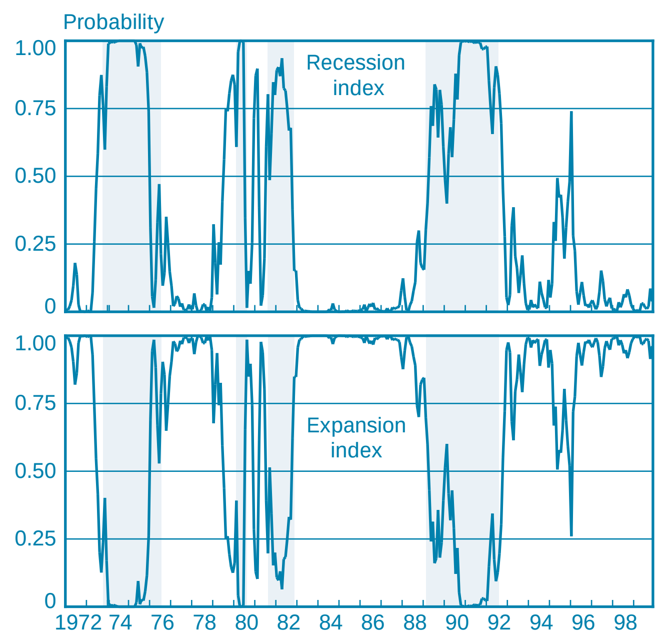
The recession and expansion indexes for New York State and New Jersey are plotted, respectively, in Charts 7 and 8, where the shading again indicates state recessions.[33] In the charts, the movements in the indexes closely correspond with the behavior of the LEI. Our recession (expansion) index typically rises prior to the advent of a contractionary (recovery) phase in state economic activity. In addition, most of the probabilities are close to either zero or one, suggesting that the indexes are arriving at a fairly strong conclusion about future changes in the direction of economic activity.

Although we do not attempt to provide a formal evaluation of the indexes and their ability to anticipate turning points, a reasonable metric to judge their performance might be based on whether the indexes indicate if the economy is more likely to be in a recession ($R_t > 0.5$) or an expansion ($E_t > 0.5$) within the next nine months. Based on this measure, the indexes generate slightly better lead times than those reported in Tables 1 and 2 for the modified rule of thumb, with no discernible change in the incidence of false signals. Taken together, these findings offer additional support for the reasonableness, reliability, and accuracy of the indexes.

Our recession (expansion) index typically rises prior to the advent of a contractionary (recovery) phase in state economic activity.

As previously discussed, the indexes offer a potential advantage over a rule of thumb by allowing for a continuous probability range and thereby provide additional information about the strength of a signal. There are several instances where this feature seems to be particularly useful in analyzing cyclical behavior in the two states. For example, the 1981 recession in New York State, unlike that of the nation, was sufficiently mild and brief that some regional analysts temper the term

Chart 7
New York State Recession and Expansion Indexes
January 1972–November 1999



Source: Authors' calculations.

Note: The shading indicates state recessions as determined by the authors.

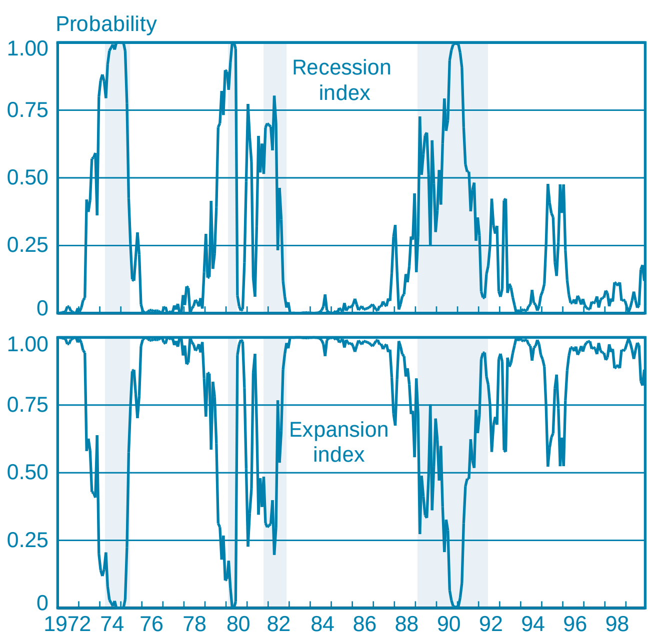
recession. This fact may explain the lack of lead time evident in the recession index for this episode. In addition, the upturn in late 1982 for New York State began very slowly before evolving into a "boom" period. In contrast, the expansion index seems to have reached a much firmer conclusion about the impending upturn.

The importance of this issue, however, seems to be most evident when considering economic developments of the mid-1990s. At the beginning of 1995, there was a pronounced slowdown of growth in the economies of New York State and New Jersey (as well as in economic activity at the national level). The recession indexes for both states seem to have anticipated this slowing correctly and display a marked and persistent rise during 1995. While a subsequent recession did not occur, there is good reason to believe that the likelihood of a recession increased during 1995. The use of a rule of thumb based only on the sign of forecasted growth rates, however, leads to a very different characterization of this period. In contrast to the indexes, both rules of thumb essentially remained silent during this period and would have provided little warning of a possible contraction in state economic activity.

Chart 8

New Jersey Recession and Expansion Indexes

January 1972–November 1999



Source: Authors' calculations.

Note: The shading indicates state recessions as determined by the authors.

Data Revisions and the Out-of-Sample Performance of the LEI Models

Two important considerations underlie the previous results. First, our indexes are constructed using models estimated over the full sample period. Second, we employ final data and do not undertake the analysis by reproducing the data that would have been available each month. In the case of real-time forecasting, however, neither of these scenarios would be relevant. For example, information arrives sequentially over a sample period. In addition, initial data releases typically are preliminary and are subject to periodic revisions. As a consequence, the within-sample results may overstate the actual forecasting performance of the indexes.

To address this concern, we now examine the out-of-sample performance of the LEI models for New York State and New Jersey. Specifically, we construct alternative LEI series for each state based on a consideration of real-time forecasting issues and then compare the series with the LEI from the within-sample analysis. Because the real-time forecasting capability of the LEI models seems to be of particular interest around turning points, we focus on the December 1986-December 1994 period. This episode provides a two-year window on either side of the cyclical peak in early 1989 and the cyclical trough in mid-1992 that occurred in each state.[34]

The construction of the alternative LEI series can be described as follows. To simulate the sequential arrival of information, we adopt an expanding sample estimation procedure. To incorporate into the analysis data revisions for the leading variables, we focus on the coincident index and substitute preliminary payroll employment data in place of the final data as one of its components. This choice is motivated by the historical decomposition in Charts 5 and 6, which suggests that past changes in the coincident index are a major contributor to the LEI, as well as our previous work (Orr, Rich, and Rosen 1999), which indicates that payroll employment is the most important component of the coincident index. To conduct the out-of-sample exercise, we estimate the LEI model for each state using the real-time analogue of the coincident index along with the other leading variables, generate a nine-month-ahead forecast, add an additional month's worth of data, and repeat the exercise.[35]

With regard to the expanding estimation procedure, it is important to recognize the nature of the restrictions that the horizon of the LEI places on the information and model estimates available to a forecaster in a real-time setting. Because we are interested in the nine-month growth rate of the CEI, the most current observation of this variable in period *t* would be

ΔC_{t-9}^{t}. From equation 1, this implies the following regression model:

(9) $$\Delta C_{t-9}^{t} = X_{t-9|t}\, \beta_{t-9} + \varepsilon_{t-9,9}\,,$$

where $X_{t-9|t}$ is a set of leading variables through month t-9 that are available in period t for the estimation of equation 9.

We can then construct a forecast of the growth in the CEI between month t and month t+9 according to the following expression:

(10) $$\Delta \hat{C}_{t}^{t+9} = X_{t}\hat{\beta}_{t-9}\,,$$

where $\Delta \hat{C}_{t}^{t+9}$ denotes the forecasted growth rate in the CEI between month t and month t+9, X_t is the corresponding value of the leading variables updated through month t, and $\hat{\beta}_{t-9}$ denotes the estimated coefficients of equation 9.

The LEI series based on the out-of-sample forecasts differs principally from its within-sample counterpart in equation 2 in terms of the model estimates. Specifically, the LEI series in equation 9 is derived from a two-step procedure that initially uses the variables $X_{t-9|t}$ as regressors to estimate the model, and then uses the variables X_t for forecasting purposes.[36] In contrast, the within-sample LEI series allows the variables X_t to serve as both the regressors and forecasting variables in the model.

The construction of the real-time analogue for the coincident index requires a timing convention for the mixture of preliminary and final payroll employment data as well as a choice of weights for the coincident indicators. Because preliminary values will correspond to the most recent observations of a time series, we assume that the data on payroll employment for a current year are preliminary and that those for the preceding years are final.[37] Following this timing convention, we can create a real-time payroll employment series for each year extending from 1986 through 1994. We can then combine the real-time payroll employment series with the other coincident indicators using the component weights from the estimation of the original coincident index model.

The outcome of this procedure is a real-time coincident index series for each year from 1986 through 1994 that can be used as a leading variable for forecasting purposes. The real-time coincident index series has the feature of incorporating preliminary payroll employment in the current year and final data in previous years. In each subsequent year, the index series would reflect not only the preliminary data for the current year, but would also incorporate the shift from preliminary payroll employment to final data for the preceding year.

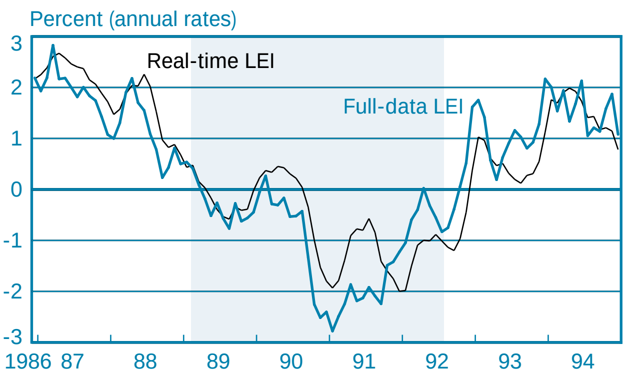
The out-of-sample forecasting performance of the LEI models for New York State and New Jersey is examined, respectively, in Charts 9 and 10. The charts illustrate the real-time LEI series from 1986 through 1994, where the LEI from the within-sample analysis is plotted again for convenience. Our principal focus is on comparing the real-time LEI series with the within-sample counterpart.[38]

As shown, the real-time LEI series appear to track closely the within-sample LEI. The indexes for New York State behave quite similarly throughout the entire episode. For New Jersey, the indexes again display similar behavior at the onset as well as during most of the 1989-92 state recession. Admittedly, there

Chart 9

New York State Real-Time LEI

December 1986–December 1994



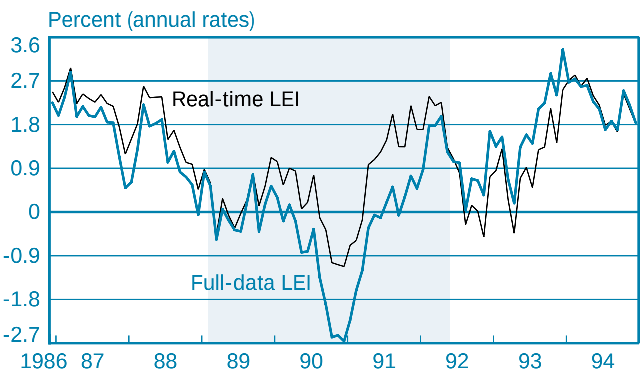
Source: Authors' calculations.

Notes: LEI is index of leading economic indicators. The shading indicates a state recession as determined by the authors.

Chart 10

New Jersey Real-Time LEI

December 1986–December 1994



Source: Authors' calculations.

Notes: LEI is index of leading economic indicators. The shading indicates a state recession as determined by the authors.

are some differences in the series during the initial phase of the subsequent recovery. Taken together, however, the evidence in Charts 9 and 10 suggests that the use of final data does not severely overstate the forecasting performance of the LEI models for New York State and New Jersey.

Conclusion

In this article, we constructed indexes of leading economic indicators for New York State and New Jersey over the 1972-99 period. The indexes are nine-month-ahead forecasts of the indexes of coincident economic indicators developed for each state in Orr, Rich, and Rosen (1999). In order to refine our forecast of future recessions and expansions, we then outlined a methodology for the construction of a recession and expansion index. These indexes provide an alternative to conventional rules of thumb by allowing the likelihood of a recession and expansion to be defined over a continuous probability range.

The historical performance of our leading indexes of future economic activity suggests that the information they convey about the timing and likelihood of a recession or expansion is quite useful and represents an improvement over the information offered by other indicators. Our results also suggest that the recession and expansion indexes provide a reliable signal of future economic turning points in New York State and New Jersey. Furthermore, the movements of our recession and expansion indexes display a close relationship with the behavior of our indexes of leading economic indicators. Accordingly, the recession and expansion indexes allow us to extend the informational content of the leading indexes by estimating the probability of an upcoming cyclical change in state economic activity.

Technical Appendix

This appendix describes the Monte Carlo procedure used to compute the recession index and expansion index for New York State and New Jersey. The recession (expansion) index is an estimate of the probability that the state economy will be in a recession (expansion) within nine months. The procedure can be described as follows.

The starting point for the analysis is the construction of a state-level index of coincident economic indicators (CEI). This series is denoted by C_t and is obtained from the estimation of a single index model developed by Stock and Watson (1989, 1991). The model is given by:

(A1)
$$\Delta Y_t = \alpha + \gamma(L)\Delta C_t + \mu_t$$
$$D(L)\mu_t = \varepsilon_t$$
$$\phi(L)\Delta C_t = \delta + \eta_t,$$

where Y_t is an (n x 1) vector of coincident variables, Δ denotes the one-period change in a variable, ε_t and η_t are serially uncorrelated disturbance terms with a diagonal covariance matrix, L is the lag operator, and $\phi(L)$, $\gamma(L)$, and (L) are, respectively, scalar, vector, and matrix lag polynomials. Because the coincident variables are assumed to be measured in either levels or log levels, the vector ΔY_t can be interpreted as differences or growth rates in the coincident variables.

Stock and Watson provide details on additional identifying restrictions and estimation of the system (equation A1) using a Kalman filter. Following Orr, Rich, and Rosen (1999), the coincident economic indicators for New York State and New Jersey are based on these four series: the monthly growth in nonfarm payroll employment, actual (and forecasted) quarterly growth in real earnings (wages and salaries), the unemployment rate, and average weekly hours worked in the manufacturing sector.

Our interest in forecasting recessions and expansions rests with the recession index and expansion index previously defined in equation 6 in the text as:

(A2)
$$R_t = Pr\,[(\Delta C_t^{t+7} < 0 \text{ and } \Delta C_t^{t+8} < 0 \text{ and } \Delta C_t^{t+9} < 0) | I_t]$$
$$E_t = Pr\,[(\Delta C_t^{t+7} > 0 \text{ and } \Delta C_t^{t+8} > 0 \text{ and } \Delta C_t^{t+9} > 0) | I_t],$$

where R_t and E_t denote, respectively, the probability of a recession and expansion in the state's economy within the next nine months, conditional on information through month t.

To evaluate the expressions in equation A2, we consider the following three-equation system previously defined in equation 7 in the text:

(A3)
$$\Delta C_t^{t+7} = X_{1t}\beta_1 + \varepsilon_{t,7}$$
$$\Delta C_t^{t+8} = X_{2t}\beta_2 + \varepsilon_{t,8}$$
$$\Delta C_t^{t+9} = X_{3t}\beta_3 + \varepsilon_{t,9},$$

where ΔC_t^{t+i} is the growth rate of the CEI between month t and month $t+i$, X_{it} denotes the vector of leading variables in the ith equation of the system available in month t, β_i is the coefficient vector of the ith equation, and $\varepsilon_t = [\varepsilon_{t,7}, \varepsilon_{t,8}, \varepsilon_{t,9}]'$ denotes the (3 x 1) vector of disturbance terms in the system. It is assumed that the disturbance vector ε_t has mean zero with covariance matrix $E[\varepsilon_t \varepsilon_t'] = \Sigma$ for all t.

Let $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$ denote the estimated parameters of the three-equation system (equation A3) and let $\hat{\Sigma}$ denote the estimated variance-covariance matrix of the observed forecast errors $(\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_{T-9}) \equiv [(\hat{\varepsilon}_{1,7}, \hat{\varepsilon}_{1,8}, \hat{\varepsilon}_{1,9}), (\hat{\varepsilon}_{2,7}, \hat{\varepsilon}_{2,8}, \hat{\varepsilon}_{2,9}), \ldots, (\hat{\varepsilon}_{T-9,7}, \hat{\varepsilon}_{T-9,8}, \hat{\varepsilon}_{T-9,9})]$. To construct the recession index in period τ, we begin with the point forecasts of the seven-, eight-, and nine-month growth rate in the CEI, which are given, respectively, by:

(A4)
$$\Delta\hat{C}_\tau^{\tau+7} = X_{1\tau}\hat{\beta}_1$$
$$\Delta\hat{C}_\tau^{\tau+8} = X_{2\tau}\hat{\beta}_2$$
$$\Delta\hat{C}_\tau^{\tau+9} = X_{3\tau}\hat{\beta}_3.$$

Next, we use simulation techniques—Monte Carlo methods—to generate a sample of artificial observations for the seven-, eight-, and nine-month growth rates of the CEI, which we denote by $[\Delta\tilde{C}_\tau^{\tau+7}, \Delta\tilde{C}_\tau^{\tau+8}, \Delta\tilde{C}_\tau^{\tau+9}]$. The set of observations for the artificial sample is generated by randomly drawing an observation from $\hat{\Sigma}$ and then adding the realization $[\tilde{\varepsilon}_\tau^{\tau+7}(1), \tilde{\varepsilon}_\tau^{\tau+8}(1), \tilde{\varepsilon}_\tau^{\tau+9}(1)]$ to the actual point forecasts based on the historical data from equation A4:

(A5)
$$\Delta\tilde{C}_\tau^{\tau+7}(1) = X_{1\tau}\hat{\beta}_1 + \tilde{\varepsilon}_\tau^{\tau+7}(1)$$
$$\Delta\tilde{C}_\tau^{\tau+8}(1) = X_{2\tau}\hat{\beta}_2 + \tilde{\varepsilon}_\tau^{\tau+8}(1)$$
$$\Delta\tilde{C}_\tau^{\tau+9}(1) = X_{3\tau}\hat{\beta}_3 + \tilde{\varepsilon}_\tau^{\tau+9}(1).$$

A full sample of artificial data for month τ is then constructed by repeating this process N times:

(A6) $$[(\Delta\tilde{C}_\tau^{\tau+7}(1), \Delta\tilde{C}_\tau^{\tau+8}(1), \Delta\tilde{C}_\tau^{\tau+9}(1)), \ldots, (\Delta\tilde{C}_\tau^{\tau+7}(N), \Delta\tilde{C}_\tau^{\tau+8}(N), \Delta\tilde{C}_\tau^{\tau+9}(N))].$$

We select $N = 1{,}000$ replications to generate the full sample of artificial data for a given month τ.

The generation of the full sample of artificial data relies on an estimate of the variance-covariance matrix ($\hat{\Sigma}$) of the disturbance terms of the system. Because of the overlapping nature of the forecasts, the off-diagonal elements of the matrix $\hat{\Sigma}$ are not expected to equal zero. For the Monte Carlo analysis, the random draws used to construct the artificial data are taken from a multivariate normal distribution.

Once a full sample of artificial data for month τ is generated, we can count the number of times the recessionary pattern occurs in equation A6:

(A7) $$R_\tau^i = 1 \; if\, (\Delta\tilde{C}_\tau^{\tau+7}(i) < 0 \; and \; \Delta\tilde{C}_\tau^{\tau+8}(i) < 0 \; and \; \Delta\tilde{C}_\tau^{\tau+9}(i) < 0)$$
$$R_\tau^i = 0 \; otherwise.$$

Similarly, we can count the number of times the expansionary pattern occurs in equation A6:

(A8) $$E_\tau^i = 1 \; if\, (\Delta\tilde{C}_\tau^{\tau+7}(i) > 0 \; and \; \Delta\tilde{C}_\tau^{\tau+8}(i) > 0 \; and \; \Delta\tilde{C}_\tau^{\tau+9}(i) > 0)$$
$$E_\tau^i = 0 \; otherwise.$$

As discussed in endnote 20, the definitions of the recessionary and expansionary patterns in equation 5 are not collectively exhaustive. Because of the dichotomous nature of recessions and expansions, however, we will exclude the indeterminate growth sequences from our subsequent calculations. Accordingly, we compute an estimate of the recession index in month τ as the ratio of the count in equation A7 to the total count from equations A7 and A8:

(A9) $$R_\tau = Pr\,[(\Delta C_\tau^{\tau+7} < 0 \; and \; \Delta C_\tau^{\tau+8} < 0 \; and \; \Delta C_\tau^{\tau+9} < 0) | I_\tau] = \frac{\left(\sum_{i=1}^{N} R_\tau^i\right)}{\left(\sum_{i=1}^{N} R_\tau^i + \sum_{i=1}^{N} E_\tau^i\right)}.$$

The expansion index in month τ is then given by:

(A10) $$E_\tau = 1 - R_\tau.$$

The recession or expansion index can then be obtained for each time period in the historical sample by selecting the relevant point forecasts of the seven-, eight-, and nine-month growth forecasts of the CEI and repeating the procedure outlined above for $\tau = 1, 2, \ldots, T$.

The specification of the three-equation system for New York State is given by:

(A11) $$\Delta C_t^{t+7} = \alpha_0 + \sum_{i=0}^{3} \beta_i \Delta C_{t-i-7}^{t-i} + \delta_0 \Delta Permits_t + \delta_3 \Delta Permits_{t-3} + \delta_6 \Delta Permits_{t-6} + \psi_0(Spread_t) + \gamma_0 \Delta LEI_t^{National} + \varepsilon_{t,7}$$
$$\Delta C_t^{t+8} = \alpha_0 + \sum_{i=0}^{2} \beta_i \Delta C_{t-i-8}^{t-i} + \delta_0 \Delta Permits_t + \delta_3 \Delta Permits_{t-3} + \delta_6 \Delta Permits_{t-6} + \psi_0(Spread_t) + \gamma_0 \Delta LEI_t^{National} + \varepsilon_{t,8}$$
$$\Delta C_t^{t+9} = \alpha_0 + \sum_{i=0}^{1} \beta_i \Delta C_{t-i-9}^{t-i} + \delta_0 \Delta Permits_t + \delta_3 \Delta Permits_{t-3} + \delta_6 \Delta Permits_{t-6} + \psi_0(Spread_t) + \gamma_0 \Delta LEI_t^{National} + \varepsilon_{t,9}.$$

The specification of the three-equation system for New Jersey is given by:

(A12) $$\Delta C_t^{t+7} = \alpha_0 + \sum_{i=0}^{3} \beta_i \Delta C_{t-i-7}^{t-i} + \delta_0 \Delta Permits_t + \delta_3 \Delta Permits_{t-3} + \delta_6 \Delta Permits_{t-6} + \psi_0(Spread)_t + \gamma_0 \Delta LEI_t^{National} + \varepsilon_{t,7}$$
$$\Delta C_t^{t+8} = \alpha_0 + \sum_{i=0}^{3} \beta_i \Delta C_{t-i-8}^{t-i} + \delta_0 \Delta Permits_t + \delta_3 \Delta Permits_{t-3} + \delta_6 \Delta Permits_{t-6} + \psi_0(Spread)_t + \gamma_0 \Delta LEI_t^{National} + \varepsilon_{t,8}$$
$$\Delta C_t^{t+9} = \alpha_0 + \sum_{i=0}^{2} \beta_i \Delta C_{t-i-9}^{t-i} + \delta_0 \Delta Permits_t + \delta_3 \Delta Permits_{t-3} + \delta_6 \Delta Permits_{t-6} + \psi_0(Spread)_t + \gamma_0 \Delta LEI_t^{National} + \varepsilon_{t,9}.$$

Data Appendix

Leading Variables

Interest Rate Spread (Percent per Annum)

Interest rates:
- One-year constant maturity securities
- Ten-year constant maturity securities

Source: Board of Governors of the Federal Reserve System.

Composite Index of Leading Economic Indicators

Average weekly hours, manufacturing
Average weekly initial claims for unemployment insurance
Manufacturers' new orders, consumer goods, and materials
Vendor performance, slower diffusion index
Manufacturers' new orders, nondefense capital goods
Building permits, new private housing units
Stock prices, S&P 500
Money supply, M2
Interest rate spread, ten-year Treasury bonds less federal funds
Index of consumer expectations
Source: Conference Board.

New York State and New Jersey

Housing Permits (Monthly, in Thousands, Not Seasonally Adjusted)

Private housing units, permit-authorized
Construction Report C40
(Seasonally adjusted by the Federal Reserve Bank of New York.)
Source: U.S. Department of Commerce, Bureau of the Census.

Coincident Variables

New York State and New Jersey

Nonfarm Payroll Employment (Monthly, in Thousands, Seasonally Adjusted)

Unemployment Rate (Monthly, Percent, Seasonally Adjusted)

Average Weekly Hours in Manufacturing (Monthly, Seasonally Adjusted)

Hours series were smoothed to remove outliers due to such factors as strikes and weather. Observations exceeding three standard deviations from trend line were adjusted.
Sources: New York State Department of Labor; New Jersey Department of Labor.

Real Earnings (Millions of Dollars, Seasonally Adjusted Annual Rate)

Wages and salaries, total
(Deflated by national consumer price index.)
Source: U.S. Department of Commerce.

Endnotes

1. For example, payroll data are a measure of a single market, and gross state product figures are available only annually and are subject to a two-year lag in their release.

2. The indexes for Pennsylvania and New Jersey are reported monthly by the Federal Reserve Bank of Philadelphia. The Texas leading index is reported bimonthly by the Federal Reserve Bank of Dallas. An index of leading economic indicators for the Massachusetts economy has recently been developed (Clayton-Matthews and Stock 1998-99).

3. The National Bureau of Economic Research (NBER) determines peaks and troughs in national economic activity. The dating of these turning points is based on the consideration of myriad variables. A major input into the dating procedure is the national index of coincident economic indicators originally constructed by the U.S. Department of Commerce (1987). The indexes are now maintained and reported by the Conference Board.

4. Crone (1998) has recently constructed an index of coincident economic indicators for each of the forty-eight contiguous states.

5. The development of state coincident indexes is constrained by the relatively short time period for which the state data are available. This constraint is particularly important because it limits the number of turning points in state economic activity that can be used to estimate and evaluate an index of leading economic indicators.

6. We discuss our specification for the New Jersey LEI in the "Empirical Results" section.

7. The Texas trade-weighted dollar is an exchange rate that measures the relative price of Texas-made products in the world market. It uses the relative shares of merchandise exports to forty-one countries to compute a (weighted-average) real value of the dollar facing Texas producers.

8. The rule currently used to determine if the national index of leading economic indicators is signaling a recession is a decline of 2 percent or more (at an annual rate) over six months coupled with a decline in the majority of the component series (Conference Board 1997).

9. The Pennsylvania LEI predicted all four of the state's recessions since 1973 with leads of five months or more and predicted two of the four recoveries; the New Jersey LEI predicted all four of the state's recessions since 1973 with leads of one to seven months, but predicted only one of the four recoveries.

10. The Texas leading index led both peaks and troughs in the coincident index by an average of two months.

11. The NBER's dating of recessions is also subject to occasional revisions.

12. As an alternative to focusing on the national composite index of leading indicators, some forecasters have used (two) consecutive quarters of negative GDP growth as the basis for predicting recessions.

13. The Stock-Watson model actually allows for the calculation of a coincident index for period t based only on currently available information ($C_{t|t}$) as well as information from the full sample ($C_{t|T}$). While the two series typically are in close agreement, we nevertheless draw a distinction between them. Specifically, we use the full-sample CEI ($C_{t|t}$) series in Chart 1 to date state business cycles. For the construction of the LEI, however, we associate the coincident index with the behavior of the $C_{t|t}$ series. This is the appropriate choice for forecasting purposes. Our technical appendix and data appendix contain, respectively, a brief description of the methodology and a list of the variables used to construct the CEI for each state.

14. We also experimented with a leading index based on a six-month forecast, but the nine-month horizon produced a slightly better lead time for some recessions and expansions. Although a longer horizon can improve the lead time, it will produce less accurate forecasts. As we will discuss shortly, the formulations of the recession index and expansion index will consider growth in the CEI over seven- and eight-month horizons as well as over a nine-month horizon. This feature of the modeling strategy therefore allows us to incorporate alternative horizons into the analysis. Because the recession and expansion indexes may be preferable for gauging the likelihood of future recessions and expansions, these considerations lessen the significance attached to our particular choice of a nine-month forecasting horizon for the LEI.

15. The VAR model essentially consists of a set of one-step-ahead forecasting equations for each variable in the system. The law of iterated projections and the estimated VAR model can be used to generate multi-step-ahead forecasts for each variable. The LEI can then be constructed from knowledge of the multi-step-ahead forecasts. See Stock and Watson (1989) and Crone and Babyak (1996) for additional discussions.

16. Ordinary least squares remains a consistent estimator of the parameters of equation 1. However, we use an estimator proposed by

Note 16 continued
Newey and West (1987) to calculate the standard errors of the model. Because the forecasting horizon of the CEI is nine months and the sampling interval of the data is one month, the Newey-West estimator allows the disturbance term to follow an eight-order moving-average (MA) process.

17. The Stock and Watson (1989, 1993) recessionary and expansionary patterns allow the monthly CEI growth sequences to be quite complex. Consequently, calculating the recession and expansion probabilities requires integrating a seventeen-dimensional normal density function.

18. Compared with the framework of Stock and Watson (1989, 1993), our strategy for forecasting recessions and expansions lacks its technical sophistication. Nevertheless, we find that our approach performs rather well. Specifically, the estimated probabilities are quite accurate in predicting both recessions *and* expansions, generally display good lead time, and generate few false signals.

19. We do not incorporate equation 3 into the analysis because the differential dating of information sets in conjunction with single-equation forecasting models does not lend itself to the type of probability calculations used to construct the recession and expansion indexes.

20. The definitions of the recessionary and expansionary patterns in equation 5 are *not* collectively exhaustive because of the existence of sequences that will not be contained in either event's set. However, their sum typically is very close to unity. As discussed in the technical appendix, we exclude the indeterminate growth sequences from the calculation of the recession and expansion indexes to ensure that their sum is equal to unity.

21. This can be seen through an examination of the system in equation 7.

22. While we restrict the discussion to a single forecast horizon, Chart 2 also provides the intuition for the methodology used to calculate the recession (and expansion) index using the system specified in equation 7.

23. This probability is also indicated by the shaded area in Chart 2 and corresponds to the likelihood of observing lower future levels of the CEI. It is worth noting that the calculation of the recession and expansion indexes described in the technical appendix assumes that the variance of the disturbance terms for the system in equation 7 is constant. Recently, McConnell and Perez Quiros (2000) have documented that a permanent decline in the volatility of U.S. GDP growth occurred in 1984. If the growth rate in the CEI for New York State and New Jersey were to experience a similar decline in volatility over our sample period, then the recession and expansion indexes would need to account for this feature of the data. Specifically, the calculation of the indexes would require different estimated variance-covariance matrices for the disturbance terms for the system in equation 7, with the matrices estimated over separate subperiods.

24. We considered other regional economic indicators such as regional consumer confidence; regional new car and light truck registrations; New York Stock Exchange member firms' profits; and employment in the chemical, manufacturing, and finance industries. However, these variables were either statistically insignificant or displayed the theoretically incorrect sign.

25. The financial market indicators consist of the value of the S&P 500 and an interest rate spread measuring the difference between the yield on ten-year Treasury bonds and the federal funds rate.

26. This delay reflects publication lags in the individual coincident indicators.

27. We used various criteria, including the Bayesian information criterion and the predictive least squares sum of squared residuals.

28. For example, the value of the LEI series for January 1980 in Charts 3 and 4 corresponds to the forecasted growth in the CEI from January to September 1980, conditional on information available through February 1980.

29. The smoothing filter is given by $s(L) = (1/6)\ (1 + 2L + 2L^2 + L^3)$, where L denotes the lag operator defined by $L^n X_t = X_{t-n}$. As described in Tables 1 and 2, the interest rate spread is also smoothed by using an average of its value during the current and previous two months.

30. One possible explanation for this result relates to the markedly different timing of the state and national recessions during this episode. Specifically, the state recessions preceded the national recessions by almost fifteen months. Thus, the state-level leading indicators by themselves may not have been fully capable of generating a downturn in the LEI prior to the onset of the recession. However, it is also important to recall that forecasting the 1990-91 recession at the national

level also proved to be problematic. For example, the leading index and the recession index developed by Stock and Watson experienced a complete breakdown and missed the 1990-91 recession in its entirety.

31. The contribution of each indicator is determined by setting all of the other components equal to zero and then calculating the relevant quantity from equation 8.

32. It is unclear how one should classify a lead time of zero months. Therefore, we do not include this event in a count of accurate leading signals.

33. The value of the recession index in January 1980 again corresponds to the probability of a recession from January to September 1980, conditional on information available through February 1980.

34. As we discuss, we employ preliminary data on state payroll employment to evaluate the out-of-sample performance of the LEI models. Because the publicly available data on this series begin in 1981, the evaluation can include only the recessionary episode of 1989-92.

35. There is the additional issue of whether the objective should be to forecast the preliminary or final value of the coincident index. The analysis will continue to specify the dependent variable of the LEI models as the nine-month growth rate in the final coincident index. Our choice is initially motivated by the belief that the preliminary payroll employment figure will act as an unbiased estimate of the final figure. More importantly, we present evidence that the LEI models work quite well in terms of their out-of-sample forecasting performance under this more stringent assumption.

36. Another difference between the two approaches is that the estimated parameter vector in equation 2 is assumed constant over the full sample period, while the estimated parameter vector in equation 9 is updated every period.

37. Annual revisions to the state payroll employment series for the previous calendar year normally are released in late February. Subsequent revisions to the payroll employment data are generally not of great importance. We thank Jason Bram for his assistance in obtaining the historical preliminary payroll employment data and constructing the real-time payroll employment series.

38. Our procedure actually results in a one-year overlap of the real-time LEI series and therefore can be used to investigate the sensitivity of the real-time forecasts to data revisions. An examination of these overlapping segments indicates that the behavior of the series was qualitatively similar. For purposes of presentation, however, the overlapping segments of the real-time LEI series are excluded from Charts 9 and 10.

References

Clayton-Matthews, Alan, and James H. Stock. 1998-99. "An Application of the Stock/Watson Index Methodology to the Massachusetts Economy." Journal of Economic and Social Measurement 25: 183-233.

Conference Board. 1997. Business Cycle Indicators. New York: Conference Board.

Crone, Theodore M. 1994. "New Indexes Track the State of the States." Federal Reserve Bank of Philadelphia Business Review, January-February: 19-31.

———. 1998. "Using State Indexes to Define Economic Regions in the U.S." Paper prepared for the Workshop on Regional Economic Indicator Models, Braga, Portugal, June 8-10.

Crone, Theodore M., and Kevin J. Babyak. 1996. "Looking Ahead: Leading Indexes for Pennsylvania and New Jersey." Federal Reserve Bank of Philadelphia Business Review, May-June: 3-14.

Kuttner, Kenneth N., and Argia M. Sbordone. 1997. "Sources of New York Employment Fluctuations." Federal Reserve Bank of New York Economic Policy Review 3, no. 1: 21-35.

McConnell, Margaret M., and Gabriel Perez Quiros. 2000. "Output Fluctuations in the United States: What Has Changed since the Early 1980s?" American Economic Review 90, no. 5 (December): 1464-76.

Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." Econometrica 55, no. 3 (May): 703-8.

Orr, James, Robert Rich, and Rae Rosen. 1999. "Coincident Indexes of Economic Activity for the New York–New Jersey Region." Federal Reserve Bank of New York Current Issues in Economics and Finance 5, no. 14.

Phillips, Keith R. 1988. "New Tools for Analyzing the Texas Economy: Indexes of Coincident and Leading Economic Indicators." Federal Reserve Bank of Dallas Economic Review, July: 1-13.

———. 1990. "The Texas Index of Leading Economic Indicators: A Revision and Further Evaluation." Federal Reserve Bank of Dallas Economic Review, July: 17-25.

Stock, James H., and Mark W. Watson. 1989. "New Indexes of Coincident and Leading Economic Indicators." In Olivier Blanchard and Stanley Fischer, eds., NBER Macroeconomics Annual: 1989, 351-94. Cambridge, Mass.: MIT Press.

———. 1991. "A Probability Model of the Coincident Economic Indicators." In Kajal Lahiri and Geoffrey Moore, eds., Leading Economic Indicators: New Approaches and Forecasting Records, 63-90. Cambridge: Cambridge University Press.

———. 1993. "A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Experience." In James H. Stock and Mark W. Watson, eds., New Research on Business Cycles, Indicators, and Forecasting, 95-153. Chicago: University of Chicago Press.

U.S. Department of Commerce. 1987. "Composite Indexes of Leading, Coincident, and Lagging Indicators." Survey of Current Business, November.

The Federal Reserve Bank of New York
Economic Policy Review

Single-copy subscriptions to the *Economic Policy Review* (ISSN 0147-6580) are free. Multiple copies are available for an annual cost of $12 for each additional subscription. Checks should be made payable in U.S. dollars to the Federal Reserve Bank of New York and sent to the Public Information Department, 33 Liberty Street, New York, NY 10045-0001 (212-720-6134). Single and multiple copies for U.S. subscribers are sent via third- and fourth-class mail. *Economic Policy Review* subscribers also receive the Bank's Annual Report.

Library of Congress Card Number: 77-646559

Note that we have changed our web site address. Please update your bookmarks.

www.newyorkfed.org/rmaghome

All Federal Reserve Bank of New York research publications—the *Economic Policy Review*, *Current Issues in Economics and Finance*, *Staff Reports*, *Research Update*, and the annual publications catalogue—are available at the Research Group's web site.

Visitors to the site can also subscribe to a free **Electronic Alert Service**. Subscribers to Electronic Alert receive timely e-mails announcing new publications as soon as they are posted on our web site. Electronic Alert also notifies subscribers of upcoming New York Fed conferences and calls for papers.