The Choice of Capital Instruments

LARRY D. WALL AND PAMELA P. PETERSON
Wall is an officer in the financial section of the Atlanta Fed's research department. Peterson is a professor in the finance department of Florida State University.
They thank James Ang, George Benston, Curt Hunter, Steve Smith, Raymond Spudeck, and other participants in workshops at Emory University and Florida State University and at the 1990 Financial Management Association and 1991 Southern Finance Association meetings for comments on a prior version of this paper.

A NUMBER OF VERY LARGE BANKING ORGANIZATIONS HAVE RECENTLY PROPOSED OR CONSUMMATED MERGERS, INCLUDING CHEMICAL AND CHASE MANHATTAN, CITICORP AND TRAVELERS, AND NATIONSBANK AND BANKAMERICA. THESE COMBINATIONS HAVE INCREASED THE IMPORTANCE OF HAVING A SYSTEM OF BANK SUPERVISION AND REGULATION THAT PROTECTS THE TAXPAYERS AND THE FINANCIAL SYSTEM WHILE AVOIDING THE IMPOSITION OF UNNECESSARY COSTS ON BANKS. THIS ARTICLE FOCUSES ON THE COSTS IMPOSED BY ONE OF THE PRIMARY TOOLS OF CURRENT BANK SUPERVISION AND REGULATION—THE MEASUREMENT AND REGULATION OF CAPITAL ADEQUACY.

Bank capital ratios have become one of the principal measures of a bank's financial condition. Capital ratios have long been an important regulatory consideration, but their importance has recently grown partly as a consequence of international efforts to harmonize bank supervisory rules and partly because of the inclusion of prompt corrective action provisions in the Federal Deposit Insurance Corporation Improvement Act (FDICIA). One important concern is that the emphasis on capital regulation will increase banks' costs and make them relatively less competitive with other financial service providers. The capital regulations may impose costs on banks to the extent that the controls reduce the subsidy value of the federal safety net; however, this increase in costs is an intended consequence of the regulations, designed to offset Federal Deposit Insurance Corporation (FDIC) liability.[1]

A more troubling question is whether the regulations impose costs that are not necessary for achieving the goals of the regulation. In particular, the current standards effectively force banks to maintain minimum levels of equity capital as measured in accounting values. Yet a variety of studies have suggested that maintaining higher equity capital levels at the cost of reduced debt levels is costly—for example, in reducing the tax shield associated with corporate interest payments. If equity is more expensive than debt, regulators should reconsider the limits they impose on substituting debt for equity.

In an earlier article in this *Economic Review*, Wall and Peterson (1996) surveyed the existing literature on banks' responses to binding capital regulations. They found that empirical evidence supports the hypothesis that capital regulations exercise a binding influence on

banks' capital positions. They then examined banks' two options for responding to binding regulations: (1) actions that increase regulatory capital ratio measures without reducing a bank's risk of failure, which Wall and Peterson call cosmetic changes, and (2) actions that increase regulatory ratios and reduce the risk of failure, which they called effective increases in capital.[2] The evidence they survey indicates that stock market participants see through cosmetic changes that artificially raise capital and that they reduce the stock price of banks engaging in such steps. Banks may effectively increase their capital through the issuance of new stock, but this action also reduces the price. One explanation of both findings is that the market may interpret cosmetic actions and new equity issuance as indicating that the bank expects weak future earnings and thus must take other steps to satisfy the regulatory requirements.

Wall and Peterson's review may be interpreted as suggesting that capital regulations impose unnecessary costs on the banking system, but this conclusion is not obvious. Financial market participants seem to be using a bank's decision to issue capital as a method of inferring the bank's future earning power. If market participants are rational and financial markets are competitive, these inferences should be correct on average, so it is not necessarily the case that capital regulations have been costly to the banking system as a whole.[3]

However, even if financial market participants make correct inferences on average, they may still be unable to fully separate banks with good prospects that have had bad luck from banks with bad prospects.[4] If market participants cannot perfectly separate the two sets of banks, then the market price of good banks may fall more than it otherwise would and the market price of bad banks may fall less than it otherwise would. In effect, the capital regulations cause good banks to issue capital at a lower price than they should, and good banks' losses are offset by bad banks' being able to issue capital at a higher price.

The market may make inferences about a bank's condition from either debt or equity issues. However, as the residual claimant on a bank's value, the value of common equity is most sensitive to market misestimation, while claims on a fixed portion of the bank's cash flow—that is, debt—are less sensitive to misestimation. Hence, capital regulations that occasionally force banks to issue new equity may impose higher costs on good banks than capital regulations that allow the bank to substitute a debt issue.

This article focuses on the question of whether existing capital regulations are imposing unnecessary costs on banks. The first section reviews the regulatory and market influences that have been hypothesized to influence banks' decisions about issuing capital. The second section describes the model used in this research. The results reported in the third section provide new evidence on the costs associated with new capital issues by banks, thereby shedding additional light on the private costs of capital regulation. The final part analyzes some reasons why regulators might choose to set minimum equity capital requirements.

Most of the private costs and benefits associated with different capital structures arise because of the differences between debt and equity.

The new evidence is obtained by analyzing the determinants of which new security, if any, a banking organization issues to meet capital regulations. Issuance of capital instruments may impose a variety of costs on banks, depending on the instrument chosen. Capital regulations have always counted common equity and at least some types of preferred stock in calculating the ratios. However, including some types of debt securities as well may enhance the ability to distinguish among the different theories of capital structure. Under the existing tier 1 risk-based and leverage capital regulations, no type of debt security is a substitute for equity; however, the capital regulations first adopted in December 1981 allowed a special type of debt called mandatory convertible debt to substitute for equity in primary capital (the equivalent of the current tier 1 capital measure). Thus, to include debt securities, the discussion looks back at the issuance decisions made under the primary capital regulations of the 1980s.

1. *Both the calculation of the capital adequacy measure and the required level have been the subject of ongoing debate. For a critical analysis of the existing rules see the Shadow Financial Regulatory Committee (1996, Statements 84, 96, 110, 112, 124, and 126), Peek and Rosengren (1997), and Jones and Mingo (1998).*
2. *An example of a cosmetic change would be selling assets that have appreciated in value but not those that have decreased in value to increase capital as measured by regulatory accounting even if the sale reduced the bank's economic capital. An example of an effective action is the issuance of new capital by a bank.*
3. *Admittedly, the conclusion that capital regulations may have not been costly might be weakened by the inclusion of risk aversion on the part of investors.*
4. *In this case the term bad luck is used as a way of referring to banks that happened to obtain an earnings draw from the lower tail of the distribution.*

The results of the empirical analysis support Wall and Peterson's conclusion that asymmetric information costs are an important part of the issuance of additional common equity. The results also suggest, though, that the option to issue debt securities as a substitute for equity may be more valuable to large banks than to smaller banks, the latter being significantly less likely to issue mandatory convertible debt.[5]

The analysis of the regulatory implications of allowing banks to substitute debt for equity in the capital structure suggests that properly structured debt is as good, or better than, equity in addressing most regulatory concerns. The area of primary regulatory concern in which equity is likely to be superior is that of minimizing the risk of failure after a bank has already incurred a loss. However, subordinated debt may be more effective in discouraging banks from taking excessive risk and therefore may reduce the probability that a bank becomes financially distressed. Moreover, even if a failure should occur the regulators retain other tools for reducing the costs to society. Thus, the one advantage of equity over debt from a regulatory perspective may not be that important.

If equity is more expensive than debt, regulators should reconsider the limits they impose on substituting debt for equity.

Theoretical Determinants of the Capital Issuance

Banks are private corporations that operate in a special regulatory environment. As private corporations, their capital structure decisions are subject to the same influences as other corporations. These influences include factors that would lead to an optimal equity-to-debt ratio in a static setting as well as dynamic adjustment costs such as the costs of issuing new equity. The regulatory environment modifies the private costs and benefits of different capital structures in two important ways. First, deposit insurance reduces the sensitivity of insured-deposit interest rates to an organization's riskiness by guaranteeing repayment even if the bank should fail. The FDIC's historical practice of extending these guarantees to other liabilities that lack de jure insurance coverage may also reduce the sensitivity of these claims to the bank's riskiness.[6] The lower sensitivity of liability rates to a bank's riskiness reduces the amount of capital shareholders would want the bank to hold for any given level of portfolio risk. The second regulatory influence is that of capital regulations. These regulations are one-sided: regulators require banks to maintain minimum levels of capital, but they virtually never object to a bank maintaining capital ratios in excess of its needs.

Regulatory Influences. The theory of security issuance for U.S. banks incorporates both the theory of capital structure for nonfinancial corporations and the unique features of banks. One of the most important features of banks is that their deposits are insured by the federal government.[7] A consequence of deposit insurance is that the cost of a large portion of a bank's funds is relatively insensitive to changes in the bank's risk, creating an incentive for banks to take greater risks. The federal government attempts to limit the exposure of its deposit insurance agency by imposing a variety of regulations on banks and by requiring banks to undergo periodic examinations.

Capital regulation is an important type of regulation. U.S. bank regulators have long been concerned with bank capital adequacy. The capital regulations during the 1970s were enforced on a case-by-case basis, successfully preventing most banks from lowering their capital ratios to a level significantly below their peers' during this period. But regulators did not prevent the industry as a whole from reducing its capital (Marcus 1983). The 1981 capital guidelines were developed to stop the reduction in capital ratios and to increase the ratios at the largest U.S. banking organizations.

The capital guidelines announced in 1981 by the Federal Reserve System for bank holding companies define two types of capital: primary capital and total capital. Primary capital includes common stock, perpetual preferred stock, retained earnings, loan-loss allowance, and mandatory convertible securities. Total capital includes primary capital plus limited-life preferred stock and subordinated debt. The standards also define three categories of bank organizations: multinational organizations (the seventeen largest bank holding companies), regional organizations (all other banks with assets in excess of $1 billion), and community organizations (those with assets of less than $1 billion).

The 1981 guidelines do not specify numerical standards for the multinational organizations, but 1981 statements expressed the regulators' expectation that these firms would increase their capital ratios. Regional organizations were expected to maintain a minimum primary capital-to-total-assets ratio of 5 percent, whereas community organizations were required to maintain a 6 percent ratio. The regulators also stated that banking organizations were generally expected to operate at capital levels above these minimal standards. The regional bank standard was extended to cover the multinational organizations in June of 1983.[8] The primary capital standard for all banking organizations was set at 5.5 percent in March 1985.

The effect of the 1981 primary capital guidelines has been to place a lower bound on the primary capital level of banking organizations. Further, the limits on the amount of mandatory convertible debt included in primary capital set a limit on the maximum total-debt-to-total-assets ratio.

Although the 1981 standards appear to have been effective in raising capital levels, the regulations also seemed to be distorting banks' portfolio decisions. In particular, the standards did not distinguish among the riskiness of different assets and also failed to explicitly incorporate off-balance-sheet exposures into the capital requirements. Subsequent to the imposition of the 1981 standards, banks were observed responding to the apparent incentives created by the capital regulations—not only were they increasing capital but they were also reducing their holdings of highly liquid, low-risk assets and increasing their exposure to off-balance-sheet contracts. In July 1988 the central banks and bank regulators of the major industrial nations reached an international agreement to implement capital guidelines that took more accurate account of the credit risks associated with banks' on- and off-balance-sheet portfolios. Interim risk-based capital standards took effect in 1990, with the full standards taking effect at the end of 1992. As a part of the risk-based capital guidelines, the narrower definition of capital excluded mandatory convertible debt, reducing its value as a substitute for equity in complying with the capital guidelines.[9] Thus, even though the primary capital standards are no longer effective, more can be learned about the relative costs of debt and equity arising from market forces by analyzing bank capital decisions under the primary capital regulations of the 1980s.

Market Influences. Market forces could potentially impose varying costs based on both the level of a bank's capital and changes in the bank's capital structure. The theoretical starting point for analyzing market forces is Modigliani and Miller's (1958) demonstration that a firm's capital structure—that is, its mix of debt and equity—does not affect its value in perfect markets. An implication of this model is that securities prices are an unbiased estimate of their intrinsic value, so the timing and type of security sold by the firm do not affect the value of the firm. Modigliani and Miller's work not only established the conditions under which capital structure is irrelevant but also told financial economists under what conditions capital structure may be relevant.[10]

Allowing banks to issue debt rather than equity may reduce their costs of complying with the capital standards.

Building on a variety of studies analyzing nonfinancial corporations' optimal capital, Orgler and Taggart (1983) develop a market model of optimal capital structure for banks.[11] In their model, the benefits to banks of lower capital ratios are more favorable tax treatment and an increase in the value of deposit insurance. The offsetting costs of lower capital ratios are the (eventual) diseconomies of scale in producing deposit services and the deadweight costs of bankruptcy that are partially borne by the bank's owners.[12] Flannery (1994) argues that agency costs also may be an important determinant of bank capital structures.[13] Lower capital ratios impose desirable

5. *One limitation of the empirical analysis is that the model has problems identifying why banks would issue preferred stock rather than mandatory convertible debt.*
6. *Although de jure deposit insurance coverage was limited to $100,000 per depositor in a domestic branch, the FDIC generally provided 100 percent coverage of all deposits and sometimes guaranteed nondeposit liabilities during the time period of this article's sample. However, the 1991 passage of FDICIA initiated a variety of steps to reduce the government subsidy to failed banks. Bank regulators appear to be generally following through on FDICIA, and deposit insurance coverage has been limited for most of the bank failures since the act's passage. However, the effectiveness of these steps in practice has not yet been fully resolved because none of the very large banks that were eligible for inclusion in this study's sample have failed since the adoption of FDICIA. See Wall (1993) for a discussion of FDICIA and its application to large banks.*
7. *Deposit insurance originates with the Federal Deposit Insurance Act, Banking Act of 1933 (48 Stat. 162 [1933]). The FDIC provides insurance for deposits, accompanied by regulatory and examining functions to monitor this insurance function.*
8. *Prior to 1983 bank holding company capital regulations were based on the Federal Reserve's general supervisory authority. In 1983 the Federal Reserve was given a specific statutory mandate by the International Lending Supervision Act of 1983 (Public Law 98-181) to require banking organizations to maintain adequate capital levels.*
9. *See Wall (1989) for a discussion of the 1981 guidelines and their replacement by the risk-based capital guidelines.*
10. *See Miller (1995) for a discussion of the relevance of the "M & M" propositions to banking.*
11. *For example, see Modigliani and Miller (1963), DeAngelo and Masulis (1980), and Masulis and Trueman (1988) on income taxes and Baxter (1967) and Kraus and Litzenberger (1973) on bankruptcy costs.*
12. *Diseconomies of scale exist if an increase in volume results in an increase in average unit costs. Deadweight losses of bankruptcy refer to costs that arise solely because of the bankruptcy and provide no social value—legal costs, for example.*
13. *See Jensen and Meckling (1976), Barnea, Haugen, and Senbet (1981), and Jensen (1986) for a discussion of agency costs in more general settings.*

limits on management and reduce the need for shareholder monitoring.[14] Conversely, lower capital increases the incentives for bank shareholders to have managers undertake riskier projects and to reject some low-risk investments. These costs of reduced capital may be mitigated, Flannery argues, by having the bank issue deposits with very short maturities so that debtholders may take effective action if the bank adopts a high-risk investment strategy. Thus, Flannery's analysis argues that banks should issue very short-term debt and maintain low capital ratios (although they would not necessarily be undercapitalized by regulatory standards).

Shrieves and Dahl (1992) and Hughes and Mester (1994) point to another agency problem that may influence banks' capital structure—managerial risk aversion. Most individuals are thought to be risk averse, and there is no good reason for thinking that bank managers are inherently more risk averse than the average shareholder. However, bank managers have proportionately far more of their total wealth (including human capital) invested in their bank than most shareholders and, as a consequence, have more to lose from the bank's failure. Thus, bank managers may choose higher capital levels than would be optimal from the shareholder's perspective. Hughes and Mester estimate bank cost functions that allow for managerial risk aversion and find support for such risk aversion.

An implicit assumption of the static trade-off models of capital structure is that the cost of adjusting a bank's capital structure is zero. Recent work that focuses on information asymmetries between managers and investors suggests, however, that the process of adjusting the capital ratio may convey important information to shareholders. An important part of the analyses of information asymmetries has focused on the issuance of new securities by corporations. Myers and Majluf (1984) examine a firm's decision to issue debt or equity and conclude that the announcement to issue equity conveys negative information to the market about the firm's value. The market may overvalue both the debt and equity of a firm. However, if the market overestimates the value of a firm, that overestimation will have a proportionately larger impact on equity because equity has the residual claim on firm's value. Thus, if management believes the intrinsic value of a firm is less than its market value, existing shareholders benefit if the firm issues equity. Otherwise, existing shareholders are best served either by the firm issuing debt or forgoing any new security issue. Prospective new shareholders realize the incentive of existing shareholders to have the firm issue new equity only if it is overvalued and, hence, interpret a new equity issue as an adverse signal about firm value. This model suggests that firms generally prefer to issue debt rather than equity. One version of this analysis holds that firms follow a pecking order in determining which securities to issue. A firm will issue debt until further debt issuance would become "excessively" costly, and then it will issue equity.

Thus, a variety of hypotheses have been offered relating to the cost and benefits of different levels of equity and changes in the equity level. Most of these costs and benefits arise from important differences between debt and equity. First, interest payments on debt receive more favorable tax treatment than dividends on equity. Second, equity may absorb losses without causing the firm to enter financial distress and bankruptcy whereas bankruptcy is often required before debtholders will accept reduced payments. This second difference has four implications: (a) higher levels of debt financing, holding other factors constant, increase the expected costs of financial distress, (b) higher levels of debt financing increase the risk to managers' human capital, (c) higher levels of debt may encourage more efficient management, and (d) higher levels of debt give equityholders an incentive to prefer a riskier investment strategy. A third difference between debt and equity is that the market is less likely to view debt issuance as an adverse signal.

One further issue is that of a possible scale effect in the cost associated with security issuance. Smaller firms (nonbanking as well as banking) are less likely to have publicly issued securities, and those having publicly issued securities are likely to have a less diverse set of types of securities. A possible explanation is that smaller security issues tend to be less liquid, in part because the costs of analyzing a security often increase at a rate that is less proportionate to the size of the issue and in part because the issue may be held by a smaller set of investors. Whatever the explanation, the implication is that the smaller banking organizations in the sample studied may be less likely to issue preferred stock or mandatory convertible debt than to expand the size of their outstanding common stock issue.

The Empirical Model

Specification. The model of security choice presented here uses multinomial logit. Roughly, the model may be thought of as simultaneously estimating linear regression models to estimate the probability that a particular type of security will be issued (see the box for specifications). In this case the concern is to explain the decision to issue one of three securities: common stock, preferred stock, or mandatory convertible debt. Because in a multinomial logit model one of the outcomes is determined by the decision made for all the other outcomes, the model requires specifying one of the possible choices as the base case and considering the probability of the other cases relative to the base case. For example, if a bank decides not to issue common or preferred stock then it must issue mandatory convertible debt. Mandatory convertible debt, the focus of this article, is the base case in the model developed below.

BOX

Estimation Technique

The choice of issuing one of the three forms of primary capital is a polychotomous, discrete decision. The decision is modeled such that the probabilities that bank i chooses to issue mandatory convertible debt (m), preferred stock (p), or common stock (c) are represented by P_{mi}, P_{pi}, and P_{ci}, respectively, such that $P_{mi} + P_{pi} + P_{ci} = 1$.

The probability of issuer i choosing one form of capital—for example, preferred stock—can be characterized as

$$P_{mi} = Pr\,[\text{Choose preferred stock}|\text{Factors affecting choice}], \quad (1)$$

or, alternatively,

$$P_{mi} = F(X_i' \beta_j), \quad (2)$$

where X is the vector of factors that influence the choice for bank i; β_j is the coefficient matrix for factors for each j alternative form of capital; Pr is the probability operator; and F is the cumulative density function.

Each equation is estimated cross-sectionally using the multinomial logit package of LIMDEP.

No generally accepted formal model incorporates all the factors discussed in the theory section to explain corporate security issuance decisions. The research reported here follows prior studies of nonbanking corporations' security issues, most notably Marsh (1982) and Jung, Kim, and Stulz (1996), in developing empirical proxies for the theoretical concepts. The discussion decomposes the security issuance decision into four parts: taxes, financial distress, security timing and pecking order, and costs related to issue size. Table 1 provides a summary of the variables and predicted signs discussed in this section.

Proxy for Taxes. Taxes may affect the capital structure decision since the issuance of a debt security, vis-à-vis an equity security, has different tax implications for the issuer. Because interest is deductible for tax purposes, the use of debt financing generally increases the value of the firm. The greater the effective tax rate (ETR), the more valuable the tax deduction and, hence, the less likely the firm is to issue either type of equity. Thus, the expected sign on the coefficients on ETR is negative in the equations for both the probability of issuing preferred stock and the probability of issuing common stock relative to mandatory convertible debt.

Proxies for Financial Distress. The relevant measure of financial distress costs for the purposes of determining optimal security issuance is the expected costs borne ex ante by the firm's existing shareholders. These costs include those borne by the firm's private creditors, given that these creditors demand a higher interest rate to compensate for higher risk levels.

The probability of distress is affected by the firm's business risk, which is in turn affected by revenue risk and operating risk. Ideally, the business risk is measured by some variable that is independent of capital structure; for example, good proxies for industrial firms would be the historical variability of sales and operating earnings. However, for banking firms whose "production" is related to the management of interest rate risk, the risk associated with operations is more complex. Since there is no comprehensive measure of the business risk of a bank that is independent of capital structure, the variability in pretax operating income (VOI) over the prior four years is used as a risk measure. The problem of variability in profitability affecting capital structure is somewhat mitigated by the fact that bank capital structures vary within narrow bounds relative to industrial corporations. The expected sign of the coefficient on VOI is positive in both the preferred and common stock equations.

Another aspect of a bank's risk is that induced by the capital structure. One proxy for a bank's financial

14. Jung, Kim, and Stulz (1996) suggest that owner/manager agency conflicts result in some nonbank firms issuing equity. Their reasoning is that some managers may want to undertake certain projects even though the projects have a negative net present value (NPV) to shareholders. Assuming these managers have effective control of the firm, their primary concern is how best to fund the negative NPV projects. If the project is funded with debt, then the expected value of the payment of interest and principal on the debt may exceed the expected returns on the negative NPV projects, eventually resulting in a shrinkage of the resources under the manager's control. However, if the projects are funded by equity, then the combination of the new equity and negative NPV projects need not reduce these resources.

The implications of this hypothesis, assuming it is true, for regulating banks' equity capital are unclear. On the one hand, capital regulation may reduce the cost to managers of issuing new equity by allowing them to claim to skeptical shareholders that a new issue is required to meet regulatory requirements. However, regulatory equity capital minimums may also increase the costs to managers if their firm should incur a substantial loss and fall into violation of the capital requirements.

TABLE 1 Explanatory Variables

Variables	Definitions	Expected Sign	
		P_p/P_m	P_c/P_m
ETR	Effective tax rate for the most recent year	–	–
VOI	Standard deviation of the ratio of pretax operating income (total assets) over the prior four years	+	+
FCR	Fixed charge coverage ratio, evaluated at the most recent fiscal year-end	–	–
BEA	Book value of common equity divided by book value of total assets	–	–
UNL	Ratio of uninsured liabilities to total assets for the most recent fiscal year-end	+	+
TBF	Binary variable that has a value of 1 if the issuer is one of the ten largest banks, 0 if otherwise	–	–
B1DUM	Binary variable that equals 1 if the ratio of the market value of equity to the book value of equity at the end of the most recent fiscal year-end is less than 1 and 0 if otherwise	–	–
B1MBK	Ratio of the market value of equity to the book value of equity at the end of the most recent fiscal year-end if the market-to-book ratio is less than 1 and 0 if otherwise	+	+
A1MBK	Ratio of the market value of equity to the book value of equity at the end of the most recent fiscal year-end if the market-to-book ratio is greater than or equal to 1 and 0 if otherwise	+	+
PCH	Price change of the common stock over the most recent fiscal year preceding the issuance announcement	+	+
CSI	Binary variable that has a value of 1 if the issuer issued common stock within the past twelve months and 0 if otherwise	–	–
PSI	Binary variable that has a value of 1 if the issuer issued preferred stock within the past twelve months and 0 if otherwise	–	+
MCI	Binary variable that has a value of 1 if the issuer issued mandatory convertible debt within the past twelve months and 0 if otherwise	+	+
LGMKT	Log of the market value of the issuer's common equity	?	–

risk based on a flow measure of its capital position is its ability to meet recurring, fixed charges for which the fixed coverage ratio (FCR) is the proxy. The greater the ability of the firm to satisfy its fixed financial obligations (that is, the greater the financial risk), the lower the probability of financial distress and the less likely the bank is to issue common or preferred stock.[15]

An alternative way to measure the risk induced by a bank's capital structure is based on its stock of capital. A measure of the capital structure is the ratio of the book value of the firm's common equity to the book value of its total assets, BEA. The expected sign on the coefficients for both types of equity issuance relative to mandatory convertible debt issuance is negative; that is, a higher existing equity ratio implies lower probability levels of issuing common or preferred stock.

Bank shareholders are concerned about that portion of bankruptcy costs that is borne by the firm's private creditors since higher bankruptcy costs imply greater risk premiums on the bank's outstanding debt. The share of the costs borne by private creditors depends in large part on the extent to which a bank's liabilities are covered by de jure or de facto deposit insurance. Thus, banks with high levels of explicit and implicit insurance are likely to face significantly lower private costs of financial distress than banks with lower levels of coverage. This study used two proxies for the level of coverage. The proxy for explicit coverage is the ratio of uninsured lia-

bilities to total assets (UNL). This ratio is expected to have a positive effect on the probability that a bank will issue either type of equity relative to the probability it will issue mandatory convertible debt. The proxy for the implicit coverage granted large banks, TBF, is a binary variable equal to one if the bank is one of the ten largest and zero otherwise. This variable is a proxy for the fact that the largest banks are considered "too big to fail" and hence are likely to have 100 percent de facto deposit insurance. The coefficients on TBF in the common and preferred equity equations are expected to have negative signs.[16]

Proxies for Security Issuance Timing and Pecking Order. One proxy for timing considerations is the ratio of the bank's market value to its book value. Bank managers claim to be reluctant to issue common equity when this ratio is less than one (Wall and Peterson 1991), saying that it results in dilution.[17] This objection is not supported by finance theory, however. The book value of a firm's stock is irrelevant to its financial decisions, and stock should be issued if the net present value of additional investments resulting from the issue exceeds the value of the stock issued. Market-to-book ratios may have greater relevance for banking because most bank assets are short-term financial assets whose market value should be close to their book value. A bank's having a market-to-book value below one suggests ex post that its management has made bad decisions, and the market may be reluctant to give these managers additional capital. Thus, banks with low market-to-book ratios may be less likely to issue new equity, especially new common stock.

Two variables are used as proxies for the effect of market-to-book ratios of less than one: a binary variable to capture any level effects of a ratio less than one and a slope term for banks with a ratio of less than one. Specifically, the binary variable B1DUM takes a value of one if the market-to-book ratio is less than one and zero otherwise, and the slope variable B1MBK takes a value of the bank's market-to-book value if the ratio is less than one. Both B1DUM and B1MBK have a value of zero for banks above the standards. Banks with ratios that are below one are less likely to issue equity, so the expected sign on B1DUM is negative in both equations. However, as B1MBK increases toward a value of one, the probability of issuing equity may increase, suggesting that the coefficients on B1MBK in both equations may be expected to have a positive sign. Banks may also take account of their market-to-book ratio in making security issuance decisions if this ratio is greater than one. Thus, the variable A1MBK is also included in the model, where A1MBK takes a value of the bank's market-to-book ratio if the ratio is greater than or equal to one and zero otherwise. The sign of the coefficients on A1MBK, like that on B1MBK, is expected to be positive.

Banks pose special problems in terms of debt maturity because a large fraction of their assets is invested in assets that either have a short maturity or are traded in liquid markets or both.

Another measure of whether a bank's stock may be perceived by a bank's managers to be over- or undervalued is the recent movement in its stock price. If management's perception of a bank's value changes more slowly than the market's, then greater levels of stock price appreciation may be associated with a higher probability that management perceives the bank's stock to be overvalued. A proxy for the recent price change in the stock is PCH, which is the price change of the common stock over the most recent fiscal year preceding the issuance announcement. The expected sign of the coefficient on PCH is positive for both types of equity.

An implication of the pecking order hypothesis is that the probability that a particular type of security is issued may be related to its own past issuance. Three dummy variables designate previous issuances within the last twelve months: CSI (issuance of common stock), PSI (issuance of preferred stock), and MCI (issuance of mandatory convertible debt). If preferred stock issue is treated as something between common stock and mandatory convertible debt, the pecking order hypothesis delivers unambiguous signs for the probability of issuing common and preferred stock relative to the probability of issuing mandatory convertible debt.[18] Under the pecking order hypothesis the probability of issuing common

15. The coverage ratio may also be interpreted as a measure of the bank's free cash flow. The cost of issuing new preferred stock or mandatory convertible debt may be reduced to the extent that it reduces the bank's free cash flow. The free-cash-flow interpretation of the coverage ratios yields the same prediction as the risk interpretation of the ratios: the probability of a firm issuing debt or preferred stock is expected to be a negative function of FCR.

16. The exact size cutoff for too-big-to-fail status is unknown and may change over time. However, the ten largest banks may be regarded as a reasonable proxy for membership in this elite group.

17. Osborn and Evans give an example of the common view that banks should not issue stock at prices below book value: "Equity issues are difficult for the money center banks since most are trading below book value" (1988, 47).

18. Preferred stock may be thought of as an intermediate case because, like debt, it commits healthy banks to making a fixed annual payment and because, like common stock, it permits the firm to suspend payments in times of severe financial distress.

TABLE 2 Sample by Type of Capital and Year of Issuance

Year	Mandatory Convertible Debt Issuance	Preferred Stock Issuance	Common Stock Issuance	Total Sample
1983	1	14	13	28
1984	29	8	10	47
1985	19	8	13	40
1986	8	9	20	37
Total	57	39	56	152

stock relative to the probability of issuing mandatory convertible debt is a positive function of MCI and PSI and a negative function of CSI. Also, the probability of issuing preferred stock is a positive function of MCI and a negative function of PSI and CSI.

Proxies for Relative Costs of Issuance. The costs of issuing new security types are hypothesized to be a decreasing function of a bank's size. Given that all banks in the sample have publicly traded common stock, this hypothesis implies that the probability of issuing preferred stock and mandatory convertible debt is an increasing function of firm size. A proxy for firm size is the market value of the firm's outstanding common equity. However, the effect of bank size on issuance cost is likely to decrease with size, implying a nonlinear relationship between the size of the firm and the cost of issuance. Thus, the natural log of banks' market value (LGMKT) is used as a proxy for the relative cost of issuance. This proxy is expected to have a negative coefficient on the probability of issuing common stock, but the expected relationship with the probability of issuing preferred stock relative to issuing mandatory convertible debt is ambiguous.

Methodology and Data. The box provides the specification of the model and a discussion of the multinomial logit. The sample consists of stock issuances from 1983 through 1986. The sample of banking organizations is taken from the banks included in the expanded annual industrial data files of Standard and Poor's Compustat.[19]

All accounting data except for mandatory convertible debt outstanding and the primary capital ratio are obtained on an annual basis from the *Bank Holding Company Financial Supplement (FR Y-9)*, collected by the Federal Reserve System, and from Compustat. Market valuation data are obtained from the University of Chicago's Center for Research in Security Prices (CRSP) data bases.

Data on the timing and amount of securities issued are obtained from Irving Trust's *Capital Securities Issued: Commercial Banking* and Lehman Brothers' *Financings by United States Banks and Bank Holding Companies since 1976* for 1983 through 1987. When Irving Trust's publication was inadequate for determining whether a debt issued qualified as mandatory convertible, the status of the security issue was also reviewed in *Moody's Banking and Finance Manual* and individual banking organizations' annual reports. The final sample consists of 152 observations. Table 2 provides a breakdown of the sample by security type and year.

Table 3 presents descriptive statistics for each of the continuous variables for each type of issuance: common stock, preferred stock, and mandatory convertible debt. The value of the market-to-book ratio (MBK) is presented rather than B1DUM, B1MKT, and A1MKT. These substitutions facilitate comparison of the average capital position and average market-to-book ratio of the three types of issuance. Note also that the mean values of the variables TBF, CSI, PSI, and MCI may be interpreted as the proportion of banks that are too big to fail, that have issued common stock, that have issued preferred stock, and that have issued mandatory convertible debt, respectively.

Several cross-sectional differences stand out in Table 3. First, mandatory convertible debt has the highest effective tax rate. Second, mandatory convertible debt is more like preferred stock than like common stock along many dimensions, including volatility of operating income, deviation from regulatory primary capital requirements, uninsured liabilities, too-big-to-fail status, and market-to-book and market value.

Estimation Results

Logit Regression. The logit estimation results appear in Table 4. The table provides the estimated coefficients and their *t*-statistics for each relative probability (relative to issuing mandatory convertible debt).

The explanatory power of the model is statistically significant, with the percentage of correctly predicted within-sample cases of 61.18. All coefficients are insignificant in the equation estimating the probability of issuing preferred stock relative to issuing mandatory convertible debt. Moreover, among the observations in which banks actually issued preferred stock, the model predicted that 43.59 percent would have issued mandatory convertible debt whereas it predicted that only 33.33 percent would have issued preferred stock.[20]

TABLE 3 Descriptive Statistics by Type of Capital[a]

Variable	Mandatory Convertible Debt Issuance	Preferred Stock Issuance	Common Stock Issuance
ETR	0.2542 (0.1364)	0.2156 (0.2725)	0.1656 (0.1297)
VOI	0.009882 (0.003098)	0.01049 (0.003353)	0.008266 (0.003658)
FCR	1.141 (0.0614)	1.1209 (0.04225)	1.158 (0.05962)
BEA	0.04699 (0.009169)	0.04842 (0.009867)	0.05285 (0.009454)
UNL	0.5158 (0.2134)	0.4863 (0.2307)	0.3593 (0.1902)
TBF	0.3684 (0.4867)	0.3077 (0.4676)	0.07142 (0.2599)
MBK	0.9462 (0.2005)	0.8842 (0.2400)	1.083 (0.2287)
PCH	0.2560 (0.2333)	0.2356 (0.1663)	0.3144 (0.1878)
CSI	0.1579 (0.3679)	0.2308 (0.4268)	0.1607 (0.3706)
PSI	0.1754 (0.3837)	0.1538 (0.3655)	0.1607 (0.3706)
MCI	0.2281 (0.4233)	0.1538 (0.3655)	0.1786 (0.3865)
MKT	1682. (1688.)	1588. (1915.)	550.7 (878.3)

[a]Mean values; standard errors are in parentheses

The sign and significance of several variables in the common stock equation are consistent with the timing of the issuance of securities. The binary variable for a bank having a market-to-book ratio of less than one (B1DUM) is marginally significantly negative (significant only at the 10 percent level), suggesting that such banks may be more likely to issue common stock. The coefficient on the market-to-book ratio of banks with a ratio of less than one (B1MBK) is significantly positive, suggesting that banks are more willing to issue common stock as their market-to-book ratio increases. Further, the coefficients on both previous preferred stock issuance (PSI) and prior mandatory convertible issuance (MCI) are positive, suggesting that banks switch to issuing common stock after exploiting opportunities to issue preferred stock and mandatory convertible debt.

The coefficient on the log of the firm's market value is significant with a negative sign, suggesting that smaller banks are more likely to issue common stock than mandatory convertible debt. This result supports the hypothesis that mandatory convertible issues are more expensive for smaller banks.

Overall, these results support prior findings in suggesting that allowing banks to issue debt rather than equity may reduce their costs of complying with the capital standards. In particular, these results support the hypothesis that allowing banks to issue debt may reduce the costs to good banks of being pooled with bad banks.

19. The data files include the primary, secondary, tertiary, full coverage, and research files. Standard & Poor's indicates that this universe contains all banks with "significant investor interest."

20. The remaining observations were predicted to have issued common stock.

TABLE 4 Multiple Logit Regression Estimation Results

Variable	Log (P_p/P_m)	Log (P_c/P_m)
ETR	1.1025 (0.5350)	0.1782 (0.9287)
VOI	85.411 (0.3856)	50.94 (0.6084)
FCR	–13.227 (0.1029)	–0.2140 (0.97691)
BEA	53.94 (0.2151)	9.949 (0.8126)
UNL	–0.6195 (0.7882)	–0.4683 (0.8315)
TBF	0.7356 (0.4735)	1.499 (0.1968)
B1DUM	2.565 (0.5207)	–7.879 (0.0548)[c]
B1MBK	0.1680 (0.9473)	10.82 (0.0019)[a]
A1MBK	2.450 (0.4047)	2.209 (0.3813)
PCH	–0.7419 (0.5538)	–0.1272 (0.9119)
CSI	0.7883 (0.2056)	0.2213 (0.7362)
PSI	0.1540 (0.83012)	2.020 (0.0120)[b]
MCI	0.06339 (0.9309)	1.5462 (0.0400)[b]
LGMKT	–0.3990 (0.3511)	–1.914 (0.00003)[a]
Constant	10.88 (0.2283)	8.700 (0.2599)
Log *L*	–128.6018	
X^2 (28)	72.55099	
Percent Predicted	61.18	

The probability that the coefficient is not equal to zero in a two-tailed t-test is shown in parentheses below the coefficient. Log *L* is the log of the likelihood at maximum, X^2 is the Chi-squared distributed statistic for the test of all nonintercept coefficients not equal to zero, and percent predicted is the percentage of correctly predicted within-sample cases, based on the largest probability using estimated coefficients.

The Chi-squared test statistic is significant at the 5 percent level of significance.

[a] indicates coefficient different from zero at the 1 percent level of significance
[b] indicates coefficient different from zero at the 5 percent level of significance
[c] indicates coefficient different from zero at the 10 percent level of significance

However, these results also suggest that the benefits of expanding the definition of capital will not accrue equally to all banks. Larger banks that can issue sufficient volumes of new securities are more likely to substitute debt for equity than are smaller banks. Further, the estimated model provides no insight into banks' choice of mandatory convertible debt versus preferred stock.

Optimal Capital Structure and Regulatory Concerns

The survey of theoretical analyses above suggests that most of the private costs and benefits associated with different capital structures arise because of the differences between debt and equity. The survey of empirical results in Wall and Peterson (1996) as well as the new results presented in the previous section suggest that a significant part of this cost takes the form of transfers from good banks to bad banks. The implication is that regulators could minimize the cost of meeting the capital guidelines to good banks by allowing banks to substitute uninsured debt for equity. The capital standards would not be costless to banks because higher capital standards would still reduce the deposit insurance subsidy to risk taking. However, the capital standards may impose little or no additional private costs to the extent that they allow firms to use debt rather than equity.

Uninsured debt is a potentially viable substitute for equity in limiting deposit insurance losses. In theory, all nondeposit liabilities became a buffer to the insurance fund with the enactment of depositor preference in 1993.[21] Further, all depositors with more than $100,000 on deposit should share any remaining losses with the FDIC under the least costly resolution provisions of FDICIA. However, capital regulations continue to focus on a limited set of equity and debt obligations.

One possible explanation for the continuing focus on equity and certain debt contracts is concern about the extent to which deposits over $100,000 and nonsubordinated liabilities would reduce FDIC losses in the event of a failure. The FDIC may, with the concurrence of the Secretary of the Treasury and the Federal Reserve Board, extend deposit insurance to deposits over $100,000. Further, nondeposit liabilities that are not contractually subordinated to deposits may be given collateral to reduce the losses on these claims should the bank fail.

Horvitz (1984) and Benston and others (1986) as well as recent speeches by Federal Reserve Governor Ferguson (1998) suggest an alternative that does not have the problems associated with depositor preference. They recommend the increased use of a type of debt called subordinated debt—debt that is junior or subordinated to all other liabilities if a bank should fail. If subordinated debt is such an easy solution, why do regulators not allow banks to substitute it for equity? Regulatory standards have in fact allowed partial substitution. Both the 1981 standards and the current standards allow subordinated debt as an element of total capital. However, both standards limit the substitution by imposing additional requirements for a narrower definition of capital that does not include ordinary subordinated debt. These requirements are the primary capital guidelines under the 1981 standards and the current tier 1 risk-based and leverage standards. Thus, the real question is why regulators do not allow unlimited substitution. Three possible objections exist to the use of subordinated debt. Two of these objections may be easily addressed within the context of the standards and their implementation. The third is more fundamental.

The first objection is that subordinated debt may not protect the FDIC. Subordinated debt does not have de jure deposit insurance coverage, but subordinated-debt holders have received de facto insurance coverage during some prior bank failures, such as that of Continental Illinois in July 1984. Flannery and Sorescu (1996) examine the extent to which subordinated obligations of banking organizations reflected the riskiness of the issuing organization between 1983 and 1991. Their findings suggest that the prices early in their sample period embed a significant probability that the FDIC would extend its coverage to include uninsured depositors.

The solution to problems posed by de facto insurance coverage is simple; however, the FDIC should not extend deposit insurance to cover subordinated liabilities. Indeed, in more recent failures the FDIC has not covered subordinated-debt holders at failed banks. Consistent with the change in FDIC policies is Flannery and Sorescu's finding that subordinated-debt holders priced individual banking organizations' default risk during the later part of their sample period.

A second objection is that the maturity structure of debt may also be important in determining banks' behavior. As noted above, Flannery argues that the maturity of a bank's debt obligations is important in minimizing conflicts between owners and creditors. Banks pose special problems in terms of debt maturity because a large fraction of their assets is invested in assets that either have a short maturity or are traded in liquid markets or both. Thus, banks are in a position to substantially change the riskiness of their investment portfolio in a matter of months or days (or perhaps even hours in a few cases). Yet the regulations for subordinated debt to be included

21. Depositor preference was passed as a part of the Omnibus Budget Reconciliation Act of 1993. Under this provision, all of the depositors at a bank, insured and uninsured, would be placed ahead of all nondeposit liability holders in the event of a bank's failure. This provision reduces the FDIC's expected losses because in the event the agency makes payments to depositors after a bank failure the FDIC assumes the same priority claim on the remaining assets as the depositors did.

in capital ratios generally require that the debt have an average maturity at issuance of at least five years.

Once again, a solution seems clear: allow or require banks to issue subordinated debt with a short maturity. Benston and others (1986) advocate that banks regularly have subordinated debt issues rolling over and that some small percentage might be redeemable. Evanoff (1992) developed a proposal in which part of the outstanding subordinated debt matures on a regular basis (such as every six months). Calomiris (1997, 1998) provides for both regular rollovers and limits on the rate the debt could pay above the riskless rate of interest. Wall (1989) developed an entire proposal he called puttable subordinated debt that would allow subordinated-debt holders to "put" their debt back to the bank, in effect simulating the discipline imposed by demand depositors in the absence of deposit insurance.[22] These various proposals for redemption of subordinated debt either at regular intervals or upon demand by subordinated creditors would allow subordinated creditors to effectively substitute debt for equity in protecting the FDIC while giving subordinated creditors a mechanism for protecting their own interests from risk-increasing strategies by equityholders.

The third possible objection to subordinated debt arises from the goal of capital requirements. If the goal of capital requirements is to protect the FDIC, then it is possible to structure subordinated obligations that will fulfill this objective. However, subordinated obligations are unlikely to help if the goal of capital requirements is to reduce the probability of failure after a bank has incurred significant losses. Subordinated debt does not provide a cushion that can absorb losses without causing failure. If the promised payments to subordinated creditors are not made in a timely manner, then the bank is illiquid and will be closed. Allowing or requiring banks to issue subordinated obligations that have a short maturity, that are partially rolled over on a regular basis, or that are puttable only increases the risk that obligated payments to subordinated-debt holders will push a weak bank into failure.

However, the argument that subordinated debt increases a bank's probability of failure after it incurs a large loss does not necessarily imply that substituting subordinated debt for equity would make the banking system less stable. Equityholders receive both the larger payout associated with risks that succeed as well as part of the losses if the gamble fails. Subordinated-debt holders cannot obtain a higher rate of return than their promised interest rate but are exposed to failed gambles. Thus, subordinated creditors are likely to provide greater incentives for banks to avoid taking excessive risks ex ante. Thus, Horvitz (1984) points out that greater reliance on subordinated debt is likely to reduce the ex ante probability that a bank will take excessive risks that would raise the probability of its failure.

Moreover, why should regulators care about the failure of an individual bank? The failure of any individual bank is not a public policy problem per se. A bank failure becomes a problem only if it causes significant losses to the FDIC or significantly reduces aggregate real (nonfinancial) economic activity. Properly structured subordinated debt protects the FDIC from losses in a manner similar to equity at insolvent banks. Moreover, a variety of studies have examined the consequences of bank failure for the real economy, and many of these studies argue that most of the adverse consequences of a bank's failure for the real economy may be offset with appropriate monetary policy.[23]

Conclusion

Banks around the world are or have been under intense regulatory pressure to raise or maintain capital levels in response to international risk-based capital guidelines. This article examines the factors that determine the type of capital banking organizations will raise by studying U.S. banks' response to the primary capital guidelines announced in December 1981.

The empirical findings suggest that asymmetric information and the costs associated with small issue size are important determinants of the security issuance decision. Bank regulators may reduce the cost of asymmetric information by allowing banks to issue qualifying debt securities to comply with all parts of the capital regulation. However, to the extent that the cost of issuing new types of securities is high, such a regulatory change may be of little value to smaller banking organizations.

Given the potential of subordinated debt to reduce the costs of regulatory compliance for at least some banks, what justification might be given for the existing focus on equity capital? This discussion considers three possible reasons: subordinated debt may not protect the FDIC, the maturity structure of debt is important in minimizing the costs of conflicts between owners and creditors, and subordinated debt is unable to reduce the probability of a bank's failure after it absorbs substantial losses. The first two objections may be easily addressed during the regulatory implementation of new rules permitting the use of subordinated debt. The third objection holds, but it ignores the role of subordinated debt in reducing the probability that a bank will incur substantial losses and the role of other mechanisms in limiting the impact of a bank's failure on the real economy.

22. Redemption of the subordinated debt in Wall's proposal is contingent on the bank remaining in compliance with the capital standards after redemption. Thus, the subordinated-debt holders could not avoid taking losses from a bank's failure merely by requesting redemption immediately prior to its failure.

23. For a survey of this literature arguing that the macro costs of bank failure need not be high, see Benston and Kaufman (1995).

REFERENCES

Barnea, Amir, Robert A. Haugen, and Lemma W. Senbet. 1981. "Market Imperfections, Agency Problems, and Capital Structure." *Financial Management* 10:7–22.

Baxter, Nevins. 1967. "Leverage, Risk of Ruin, and the Cost of Capital." *Journal of Finance* 22:395–403.

Benston, George J., Robert A. Eisenbeis, Paul M. Horvitz, Edward J. Kane, and George G. Kaufman. 1986. *Perspectives on Safe and Sound Banking.* Cambridge, Mass.: MIT Press.

Benston, George J., and George G. Kaufman. 1995. "Is the Banking and Payments System Fragile?" *Journal of Financial Services Research* 9:209–40.

Calomiris, Charles W. 1997. *The Postmodern Bank Safety Net: Lessons from Developed and Developing Economies.* Washington, D.C.: AEI Press.

———. 1998. "Building an Incentive Compatible Safety Net: Special Problems for Developing Countries." *Journal of Banking and Finance* (forthcoming).

DeAngelo, Harry, and Ronald W. Masulis. 1980. "Leverage and Dividend Irrelevancy under Corporate and Personal Taxation." *Journal of Finance* 35:453–64.

Evanoff, Douglas D. 1992. "Preferred Sources of Market Discipline: Depositors vs. Subordinated Debt Holders." Federal Reserve Bank of Chicago Working Paper No. 21.

Ferguson, Roger W., Jr. 1998. "The Changing Banking Environment and Emerging Questions for Public Policy." Remarks before the Ohio Bankers Day Conference, Columbus, Ohio, March 19.

Flannery, Mark J. 1994. "Debt Maturity and the Deadweight Cost of Leverage: Optimally Financing Banking Firms." *American Economic Review* 84:320–31.

Flannery, Mark J., and Sorin M. Sorescu. 1996. "Evidence of Bank Market Discipline in Subordinated Debenture Yields: 1983–1991." *Journal of Finance* 54:1347–77.

Horvitz, Paul M. 1984. "Subordinated Debt Is Key to New Bank Capital Requirement." *American Banker,* December 31.

Hughes, Joseph P., and Loretta J. Mester. 1994. "Bank Managers' Objectives." Federal Reserve Bank of Philadelphia Working Paper 94-8, April.

Jensen, Michael C. 1986. "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers." *American Economic Review* 76:323–29.

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

Jones, David, and John Mingo. 1998. "Industry Practices in Credit Risk Modeling and Internal Capital Allocations: Implications for a Models-based Regulatory Capital Standard." Paper presented at conference, Financial Services at the Crossroads, Federal Reserve Bank of New York, February 26–27.

Jung, Koolyul, Yong-Cheol Kim, and René M. Stulz. 1996. "Timing, Investment Opportunities, Managerial Discretion, and the Security Issue Decision." *Journal of Financial Economics* 42:159–85.

Kraus, Alan, and Robert Litzenberger. 1973. "A State-Preference Model of Optimal Financial Leverage." *Journal of Finance* 28:911–22.

Marcus, A.J. 1983. "The Bank Capital Decision: A Time Series Cross Analysis." *Journal of Finance* 38:1217–32.

Marsh, Paul. 1982. "The Choice between Equity and Debt: An Empirical Study." *Journal of Finance* 37:121–44.

Masulis, Ronald W., and Brett Trueman. 1988. "Corporate Investment and Dividend Decisions under Differential Personal Taxation." *Journal of Financial and Quantitative Analysis* 23:369–83.

Miller, Merton H. 1995. "Do the M&M Propositions Apply to Banks?" *Journal of Banking and Finance* 19 (June): 439–89.

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

———. 1963. Corporate Income Taxes and the Cost of Capital." *American Economic Review* 53:433–43.

Myers, Stewart C., and Nicholas S. Majluf. 1984. "Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have." *Journal of Financial Economics* 13:187–221.

Orgler, Yair E., and Robert A. Taggart Jr. 1983. "Implications of Corporate Capital Structure Theory for Banking Institutions: A Note." *Journal of Money, Credit, and Banking* 15:212–21.

Osborn, Neil, and Garry Evans. 1988. "Cooke's Medicine: Kill or Cure?" *Euromoney* (July): 35–54.

Peek, Joe, and Eric S. Rosengren. 1997. "How Well Capitalized Are Well-Capitalized Banks?" Federal Reserve Bank of Boston *New England Economic Review* (September/October): 40–50.

Shadow Financial Regulatory Committee. 1996. "Shadow Financial Regulatory Committee Statements Numbers 70–134." December supplement in *Journal of Financial Services Research* 110, edited by George G. Kaufman.

Shrieves, Ronald E., and Drew Dahl. 1992. "The Relationship between Risk and Capital in Commercial Banks." *Journal of Banking and Finance* 16:439–57.

Wall, Larry D. 1989. "Capital Requirements for Banks: A Look at the 1981 and 1988 Standards." Federal Reserve Bank of Atlanta *Economic Review* 74 (March/April): 14–29.

———. 1993. "Too Big to Fail after FDICIA." Federal Reserve Bank of Atlanta *Economic Review* 78 (January/February): 1–14.

Wall, Larry D., and Pamela P. Peterson. 1991. "Valuation Effects of New Capital Issues by Large Bank Holding Companies." *Journal of Financial Services Research* 5:77–87.

———. 1996. "Banks' Responses to Binding Regulatory Capital Requirements." Federal Reserve Bank of Atlanta *Economic Review* 81 (March/April): 1–17.

Why Minimum Wage Hikes May Not Reduce Employment

MADELINE ZAVODNY
The author is an economist in the regional section of the Atlanta Fed's research department. *She thanks Tom Cunningham and Mary Rosenbaum for helpful suggestions. She also thanks, without implicating, Steve Pischke, on whose lectures much of this article is based.*

ECONOMISTS HAVE TRADITIONALLY AGREED THAT INCREASES IN THE MINIMUM WAGE HAVE ADVERSE EFFECTS ON EMPLOYMENT, PARTICULARLY AMONG YOUNG AND UNSKILLED WORKERS.[1] SEVERAL STUDIES HAVE CHALLENGED THIS CONVENTIONAL WISDOM, HOWEVER, FINDING THAT MINIMUM WAGE INCREASES DO NOT APPEAR TO LOWER EMPLOYMENT AMONG TEENS AND RETAIL WORKERS. THESE STUDIES HAVE INFLUENCED THE DEBATE ON WHETHER TO RAISE THE MINIMUM WAGE; PRESIDENT CLINTON, FOR EXAMPLE, STATED IN 1996 THAT "STUDIES SHOW THAT A MODERATE INCREASE IN THE MINIMUM WAGE, ESPECIALLY IN A STRONG ECONOMY, DOES NOT INCREASE THE UNEMPLOYMENT RATE" (PRESIDENT 1996).

Policymakers who support raising the wage floor may accept findings that minimum wage increases do not adversely affect employment, but economists question them. Critics have charged that the studies use poor data or incorrect methodologies. Some economists also argue that no convincing theoretical model predicts that minimum wage increases do not reduce employment. Many of these economists believe that low-wage labor markets are instead characterized by a competitive model that predicts that an increase in the wage floor should always reduce employment. This article describes and evaluates several alternative models that may explain the controversial recent findings and proposes avenues for future research that would help determine the validity of these models.

Recent Research

Recent studies finding that higher minimum wages do not result in lower employment have used a variety of data and methods. Several of them use long time series of data to estimate the relationship between the level of the minimum wage and the proportion of teens who are employed.[2] Other research compares employment levels before and after a specific minimum wage increase. These "difference-in-differences" studies compare the changes in employment among groups strongly affected by a minimum wage hike and relatively unaffected groups.

Several time-series studies of minimum wage effects on teen employment rates do not find that higher minimum wages are associated with significantly lower

employment rates (Neumark and Wascher 1995a; Card, Katz, and Krueger 1994; Wellington 1991). These studies use an econometric method called regression to estimate the effect of the minimum wage on the teen employment-to-population ratio. Wellington (1991) uses national data from 1954 to 1986, and the other two studies use state-level data from the 1970s and 1980s. They generally find a small negative correlation between teen employment and the minimum wage that is not statistically different from zero. Their results are in marked contrast to earlier studies summarized by Brown, Gilroy, and Kohen (1982), which typically concluded from similar methodologies that teen employment rates fell by at least 1 percent when the minimum wage rose by 10 percent. Wellington indicates that the difference in the results is due to including data from the 1980s, a period when the real (inflation-adjusted) minimum wage fell. Card and Krueger (1995), however, suggest that methodological problems biased the results in earlier studies.

Several studies that use a difference-in-differences method also find that minimum wage hikes do not significantly reduce employment. These studies compare the changes in employment between two groups, only one of which is strongly affected by the increase. Card (1992a) compares the effect of the 1990 federal minimum wage increase on teen employment in high-wage states that have a low fraction of teen workers earning less than the new minimum wage with its effect in low-wage states that have a high proportion of affected teens. Standard theory predicts that employment should fall relatively more in low-wage states, but the results indicate similar employment changes in low- and high-wage states. Card (1992b) finds that employment among teens and in retail trade in California did not fall relative to employment in other places after a $0.90 increase in the state's minimum wage in 1988.[3] Katz and Krueger (1992) examine the effects of the 1990 and 1991 federal minimum wage increases on fast-food restaurants in Texas. They find that employment growth was similar at establishments that had to raise their wages to comply with the laws and at higher-paying fast-food restaurants. Card and Krueger (1994, 1998) report that employment at fast-food restaurants in New Jersey did not decline relative to levels in neighboring Pennsylvania when New Jersey raised its minimum wage by $0.80 in 1992.

These findings appear to contradict the predicted negative effect of minimum wage increases on employment in a competitive labor market, leading some economists to question the applicability of the competitive model. Before examining the validity of alternative models, a description of the competitive model is necessary. The next section describes the effect in a simple competitive model of imposing a minimum wage, or raising an existing minimum wage, on employment. Later sections present modifications of the model to examine under what circumstances a higher minimum wage would not reduce employment and discuss whether empirical evidence supports use of these models.

Recent research has challenged the conventional wisdom among economists that increases in minimum wages lower employment among low-wage workers.

The Basic Competitive Model

The simplest competitive model posits a labor market with many identical firms and homogeneous workers. The model assumes that workers must be paid more to induce them to supply additional labor and that the value of the last unit of labor, or marginal product of labor, declines as labor increases. These assumptions generate an upward-sloping labor supply curve and a downward-sloping labor demand curve that together determine the market-clearing equilibrium wage. Although the market labor supply curve is positively sloped, individual firms are assumed to face a horizontal, or perfectly elastic, labor supply curve. That is, each firm is small enough that it can hire workers without affecting the equilibrium wage. The market labor demand curve is the horizontal sum of the individual firms' labor demand curves.

In this simple model, labor is the only input used in production. The number of units each firm can produce is given by a production function $f(l)$, where l is the number of workers employed. Firms pay each worker the market wage, w, and sell their output at a constant price, p, per unit. Firms choose the quantity of labor that maximizes

1. *In one survey, 90 percent of U.S. economists agreed that an increase in the wage floor would increase unemployment among young and unskilled workers (Frey and others 1984). A recent survey asked labor economists the expected percentage change in teen employment if the minimum wage were increased 10 percent (Fuchs, Krueger, and Poterba 1997). The mean response was that teen employment would fall by 2 percent, and the median response was a 1 percent decline.*
2. *Many studies of the effect of minimum wages focus on teens because a relatively high proportion of teens is affected by minimum wage increases. However, not all teens are low-wage workers, so these studies underestimate the effect of a minimum wage increase on affected workers.*
3. *Kim and Taylor (1995) find that employment in retail trade in California grew more slowly in industries and counties that experienced larger wage increases after the minimum wage hike. Card and Krueger (1995) challenge Kim and Taylor's findings.*

profits, $\pi(l)$. Each firm's profit maximization problem is then

$$\max \pi(l) = pf(l) - wl. \quad (1)$$

Each firm maximizes profits by setting the value of the marginal product of labor equal to the marginal cost of labor. Algebraically, the marginal product of labor is the derivative of $f(l)$ with respect to l, or $\partial f(l)/\partial l$. The value of this marginal product is the price of a unit of output multiplied by the marginal product of labor, and the marginal cost of labor is the wage. The profit-maximizing condition is then

$$p \times \partial f(l)/\partial l = w. \quad (2)$$

Equation (2) gives each firm's optimal quantity of labor in terms of the price and the wage. The total quantity of labor employed is simply the number of firms times the amount of labor given by equation (2) since all firms are assumed to be identical.

In the basic competitive model, imposing a binding minimum wage reduces employment. Suppose a minimum wage above w is imposed. Since price is assumed to be constant and the marginal product of labor rises as the quantity of labor hired declines, the only way a firm can satisfy equation (2) is by reducing employment. The magnitude of the employment effect depends on the slope of the labor demand curve. If the demand for labor changes little as the wage changes, or, in other words, if the market demand curve is steeply sloped, then firms' demand for labor is inelastic and the level of employment will be fairly unresponsive to changes in the wage. Conversely, the more elastic each firm's demand for labor is and the flatter the market labor demand curve is, the more employment will fall when the wage increases. Similarly, if a wage floor already exists, raising it must lower employment. This model is inconsistent with research that finds that minimum wage increases do not reduce employment since it predicts that a binding minimum wage always lowers employment by some amount, other things being equal.

The basic competitive model makes many simplifying assumptions that are unlikely to be true, even in low-wage labor markets. The model assumes that all workers have the same skill level and that the output price does not adjust even though the imposition of a minimum wage causes firms to reduce their output. The competitive model also assumes that firms can hire an unlimited number of workers at the market wage instead of having to offer a higher wage in order to attract more workers. The minimum wage is exogenously imposed in the model, so its level is assumed not to depend on the expected effect on employment. These simplifications are modified in the models discussed in the following sections. All of the modified models predict that imposing a minimum wage can reduce employment or have no effect, but some also predict that a minimum wage can increase employment.

Alternative Models

Substitution. Incorporating workers with different skill levels in the basic competitive model yields the prediction that a minimum wage will lower employment among low-wage workers but may not lower total employment. Suppose that there are two types of workers, skilled and unskilled, and firms can imperfectly substitute among the two. The market-clearing wage for unskilled workers is w_1, and the market wage for skilled workers is w_2. Each firm's profit maximization problem is then

$$\max \pi(l_1, l_2) = pf(l_1, l_2) - w_1 l_1 - w_2 l_2, \quad (3)$$

which is similar to equation (1) but now has two types of labor.

Profit maximization requires that the ratio of the value of the marginal products of the two types of workers be equal to the ratio of their wages, or

$$\frac{\partial f(l_1, l_2)/\partial l_1}{\partial f(l_1, l_2)/\partial l_2} = \frac{w_1}{w_2}. \quad (4)$$

Equation (4) yields the ratio of skilled to unskilled workers hired at each firm.

If a minimum wage above the market-clearing wage of unskilled workers but below the wage of skilled workers is imposed, the ratio of skilled to unskilled workers will rise. Firms can either reduce the number of unskilled workers and leave the number of skilled workers unchanged or substitute skilled for unskilled workers. Which outcome occurs depends on the specific form of the production function. If all firms hire more skilled workers, the market wage for skilled workers, w_2, is likely to rise, and such a wage increase will dampen the increase in the number of skilled workers employed. Employment of workers who initially earn less than the minimum wage declines and the employment of higher-paid workers may rise, but the effect on total employment is indeterminate. Under certain assumptions, the negative effect on employment of unskilled workers outweighs any positive effect on skilled workers.[4] The total employment effect cannot be positive in this model because the wages of at least one, and possibly both, types of workers increase. Similarly, employment of unskilled workers will fall, and the effect on total employment is indeterminate without further assumptions if an existing wage floor is raised to some level between w_1 and w_2.

Empirical evidence provides indirect support for the substitution model. Individuals who have lower wages or are likely to be less skilled appear to be more adversely affected by minimum wage increases than other workers. Currie and Fallick (1996) and Neumark and Wascher (1995b) find that teens who initially earned less than a

CHART 1
Individual Firm's Response to Wage Increase

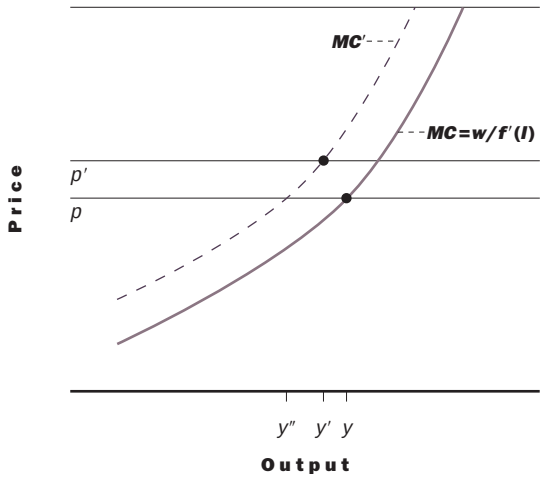
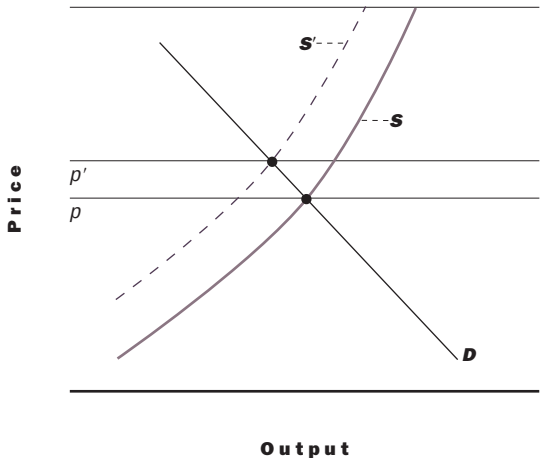


CHART 2
Industry Response to Wage Increase



subsequently imposed minimum wage are less likely to remain employed a year later than teens who initially earned more than the new minimum wage. Katz and Krueger (1992) suggest that minimum wage increases caused Texas fast-food restaurants to substitute full-time for part-time employees, who may be less skilled. Research has also documented a "ripple effect," in which the wages of workers who earn more than the minimum wage increase when the minimum wage is raised (Card and Krueger 1995). This finding is consistent with the fact that demand for higher-skilled workers increases when the wages of lower-skilled workers are forced up by the minimum wage.

There is little evidence, however, directly indicating that substituting higher-skilled workers for lower-skilled workers when the minimum wage rises leaves total employment unaffected. Currie and Fallick (1996) and Neumark and Wascher (1995b) both suggest that minimum wage increases reduce employment among low-wage teens, who presumably have low skill levels. However, total teen employment may still be unaffected by minimum wage hikes if higher-skilled teens replace lower-skilled teens. Indeed, results reported by Neumark and Wascher suggest that eighteen- and nineteen-year-olds and white teens displace sixteen- and seventeen-year-olds and minority teens from their jobs when the minimum wage increases. Further research using individual-level data is needed to confirm that there are positive employment effects among more highly skilled teens and that this substitution accounts for the failure to find overall negative effects. Research on whether firms substitute among workers of different skill levels and whether their doing so causes total employment to remain unchanged is also needed. The paucity of firm-level data and the lack of clear measures of skill levels limit research on whether firms substitute among workers when the minimum wage rises. Case studies and interviews with managers would advance the understanding of the effects of minimum wage increases.

Price Effects. The basic competitive model with one or two types of workers, outlined above, assumes that prices do not change when the minimum wage increases and firms reduce employment. However, the decline in employment is likely to cause total output to fall, and, in turn, the decrease in output will normally cause prices to rise. An increase in prices ameliorates the decline in employment but is unlikely to completely counteract it, depending on the extent to which supply and demand respond to price changes.

The effect of allowing price to change when a wage floor is imposed (or raised) is illustrated in Charts 1 and 2. Chart 1 shows the supply curve of an individual firm, where the firm's supply curve is the same as its marginal cost curve. Imposing a binding minimum wage will cause the firm's marginal cost to increase in proportion to the increase in the wage; that is, the firm's marginal cost curve shifts from MC to MC'. Chart 2 shows both the industry supply curve, which is the horizontal sum of the individual firms' supply curves, and the demand curve faced by the industry. As each firm's marginal cost increases, the industry supply curve shifts from S to S'. The decline in industry output causes the price to increase from p to p'. If the price did not increase, each firm's output would fall from y to y'' in Chart 1, but the rise in the price causes output

4. For example, Card and Krueger (1995) demonstrate that if a constant-returns-to-scale production function is assumed, total employment falls if there is also a third, nonlabor input.

to fall only from y to y'. The magnitude of the change in price will depend on the elasticities of the supply and demand curves. If demand is completely inelastic, or the demand curve is vertical in Chart 2, employment does not fall because the price increase completely offsets the tendency for the wage increase to reduce employment.

Two studies find that restaurant prices tend to rise when minimum wages increase. Research on price effects of minimum wage hikes has focused on the restaurant industry because of the prevalence of low-wage workers in the industry. Card and Krueger (1995) examine the correlations between a price index of the cost of food eaten away from home or the price of a hamburger and the fraction of restaurant workers in a city affected by the federal minimum wage increases in 1990 and 1991. They find that prices appear to have risen more quickly in cities that had a higher fraction of affected workers. Aaronson (1997) also finds that restaurant prices increase when the minimum wage rises. Both studies suggest that price increases are approximately equal to the increase in labor costs due to the minimum wage increase.

Several models can explain why employment does not appear to fall when the minimum wage increases, but further research is needed to determine their validity.

However, there is also some evidence that prices do not adjust. Katz and Krueger (1992) find no evidence of relative price increases at fast-food restaurants in Texas that were more affected by a minimum age increase. Card (1992a) finds that fast-food prices and a food-away-from-home price index rose at similar rates in California and in comparison areas after California raised its minimum wage. Even if prices rise when the minimum wage increases, price effects are unlikely to fully offset disemployment effects since demand in the restaurant industry does not appear to be completely inelastic.[5] The demand for labor would have to be unresponsive to wage changes for price effects to completely counteract the disemployment effects of a minimum wage increase, and it is not.

The "hungry teenager" theory offers another explanation for the failure to find negative employment effects in some studies (Kennan 1995). A minimum wage increase is likely to boost the earnings of some workers, and these workers may spend their extra income on low-wage goods and services, such as fast food. In effect, as one restaurant owner stated, "Our employees are our customers."[6] Such an increase in demand could offset the disemployment effect of the minimum wage hike if the demand curve shifts out enough to counteract the inward shift of the supply curve in Chart 2. Research indicates that minimum wage hikes do raise the earnings of low-wage workers on average (Neumark and Wascher 1997), but there is no direct evidence that this extra income is spent in low-wage industries.

Monopsony. Another model that almost invariably appears in textbook discussions on the minimum wage is the traditional monopsony model, which predicts that minimum wages can raise employment over a limited range. A monopsonist is a firm that faces an upward-sloping labor supply curve; a firm that is a perfect competitor in the labor market faces a horizontal labor supply curve and can hire an unlimited number of workers at the market-clearing wage. A monopsonistic firm, in contrast, must raise the wage it offers in order to hire additional workers. The monopsonist has to pay $w(l)$ to each worker to hire l workers, and $\partial w(l)/\partial l$ is positive because the labor supply curve for the firm is positively sloped. If workers are homogeneous and the output price is constant, the firm's profit-maximization problem is

$$\max \pi(l) = pf(l) - w(l)l. \qquad (5)$$

The monopsonist determines the quantity of labor to hire by setting the value of the marginal product equal to the marginal cost of labor, or

$$p \times \partial f(l)/\partial l = w(l)\left[1 + \frac{l \times \partial w(l)/\partial l}{w(l)}\right]. \qquad (6)$$

The marginal cost of labor is no longer equal to the wage, as in equation (2). Instead, the cost of hiring an additional worker is the wage paid to that worker plus the increase in the wages of all current workers. The monopsonist determines the quantity of labor to hire by setting the value of the marginal product equal to the marginal cost, as given by equation (6). The wage paid to each worker is determined from the labor supply schedule $w(l)$, as shown in Chart 3. The equilibrium outcome is given by w and l in Chart 3.

The gap between the marginal cost of labor and the wage allows a minimum wage to potentially increase employment. Suppose a minimum wage of w' is imposed in Chart 3. The firm has to raise the wages of current workers to w' and can also hire some additional workers at the minimum wage; the firm increases employment to l' because the value of the marginal product exceeds the marginal cost of labor for up to l' workers. In Chart 3 the minimum wage acts as the marginal cost curve up to the point where the wage floor intersects the labor supply curve; the dashed horizontal line at w' effectively replaces the marginal cost curve MC up to l'. Beyond that point, the marginal cost curve reverts to the original, upward-sloping curve given by MC. It is unprofitable for the firm to hire more than l' workers because the cost of hiring each additional worker beyond l' exceeds the value of that worker's marginal product.

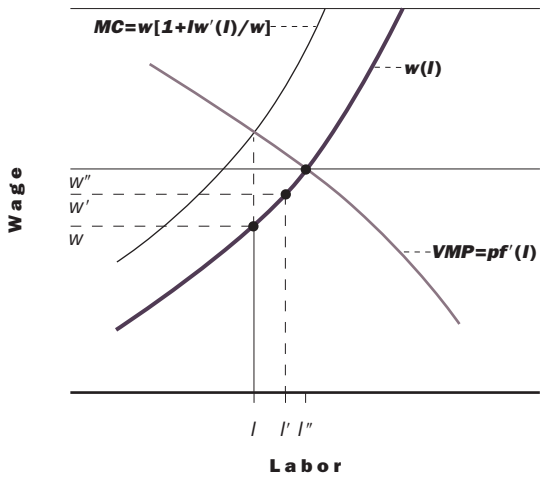
A minimum wage of w'' maximizes employment in Chart 3. Setting the minimum wage at the point at which the labor supply schedule and the value of the marginal product are equal replicates the competitive model's outcome. If the minimum wage rises above w'', employment falls because the monopsonist now determines employment from the intersection of the wage floor and the value of the marginal product curve. Imposition of a minimum wage or an increase in an existing wage floor can thus cause employment at a monopsonistic firm to rise, fall, or remain constant, depending on the level of the minimum wage.

Many economists doubt that the monopsony model explains recent research on the effects of minimum wage increases.[7] The model is generally believed to apply to firms that hire a large proportion of the workers in a labor market and seems unlikely to describe the low-wage labor market, which is usually characterized by a large number of small firms. In particular, it seems unlikely to apply to the fast-food restaurants that Card, Katz, and Krueger focus on in several studies. In addition, research on prices is inconsistent with the monopsony model's result that a minimum wage hike raises employment. If a minimum wage increase raises total employment, output should also increase. An increase in output should lower prices, and there is no clear evidence that prices fall as the minimum wage increases.

Another difficulty with the monopsony model is that its predictions for an industry may differ from its predictions for a firm. The traditional monopsony model is designed to describe a single firm that has power in the labor market, not an industry. Total employment may not increase after a minimum wage is imposed even if every firm in a labor market is a monopsonist that experiences an increase in employment. Indeed, total employment must fall in the model if each monopsonistic firm in an industry earns zero profits before a minimum wage is imposed. Suppose each firm faces a fixed cost equal to the difference between the value of the marginal product and the wage w in Chart 3. If a minimum wage is imposed, the cost of labor and the fixed cost exceed a firm's revenue. The only way a firm can return to zero losses at the higher employment level is for the price to increase. The industry price increases only if total output falls, as it does if total employment falls. If each firm in a perfectly competitive industry has monopsony power in the labor market, imposing a minimum wage raises employment at each remaining firm but lowers total employment because some firms leave the industry.

One low-wage industry that might be characterized by monopsony is restaurants with tipped workers. Wessels

CHART 3
Monopsonistic Firm's Response to Wage Floor



(1997) posits that if a restaurant hires an additional worker, average tips per worker fall and the restaurant has to make up the loss in each worker's wages. This dynamic creates the gap between the marginal cost of labor and the wage that characterizes monopsony. Wessels also shows that employment in the restaurant industry first rises and then falls as the minimum wage for tipped workers increases, matching the prediction of the monopsony model. The possibility of monopsonistic effects among tipped workers may explain findings that minimum wage increases do not reduce employment among teens since a large portion of teens work in the restaurant industry; positive effects among some tipped workers may counteract employment losses in other industries. Further research comparing employment effects among tipped workers and other workers is needed before accepting this theory as the explanation for recent findings. In addition, Wessels's theory cannot explain the recent findings on fast-food establishments, which do not have tipped workers.

Dynamic Monopsony. Some models that focus on firms' search for workers and workers' search for jobs imply that a minimum wage increase may not reduce employment. These search models are like the traditional monopsony model in that firms set wages instead of acting as price takers in the labor market, but they are considerably more complex. Firms' ability to hire or retain workers depends on their wages and on the wages offered by other firms. Firms that offer relatively higher wages attract and keep more workers, and if workers are of varying quality, firms that offer higher wages have higher-quality workers.

5. D. Brown (1990) estimates that the elasticity of demand for the restaurant industry is –0.2 and the elasticity of demand for the fast-food industry is –1. Demand is completely inelastic if its elasticity is zero, in which case the demand curve is vertical.

6. See Wall Street Journal, *November 20, 1996.*

7. For further discussion, see, for example, C. Brown (1995), Hamermesh (1995), and Welch (1995).

Imposing a minimum wage or raising an existing wage floor can raise the wages offered by some firms and, under certain assumptions, will not lower employment at those firms.

The dynamic monopsony model most commonly cited in the minimum wage literature is Burdett and Mortensen (1989).[8] In the model, employed individuals accept any job offer that exceeds their current wage, and unemployed individuals accept any offer that exceeds their reservation wage (the lowest wage at which they are willing to work). Under the assumption that firms are identical and workers are equally productive, Burdett and Mortensen show that there is a distribution of wages across firms, with larger firms paying higher wages. All firms earn equal profits. If all unemployed workers have the same reservation wage, the lowest wage offered is the reservation wage. A minimum wage above the reservation wage has no effect on employment in the model because all offers were already acceptable to unemployed workers. A minimum wage merely transfers profits from firms to workers. If reservation wages differ across workers, a minimum wage can raise employment because it increases the likelihood that unemployed workers receive offers above their reservation wages. A wage floor can lower employment, however, if it is set too high relative to workers' productivity.

Dynamic monopsony, or imperfect search, models seem more applicable than the traditional monopsony model to low-wage labor markets. As Machin and Manning (1992) explain, a key feature of dynamic monopsony models is that workers have imperfect information about the job opportunities offered by different firms. Such models often assume that workers' knowledge about the wage distribution is limited to the offers they receive. An employer that offers a lower wage than other firms loses workers slowly over time as workers receive better offers. In the perfect competition model, in contrast, workers have perfect information, and a firm that offers a wage below the market level will lose all of its workers instantly. The dynamic monopsony model seems relevant for markets with a large number of small employers, which characterizes many low-wage industries.

Search models are also appealing to many economists because they can explain many empirical regularities observed in the labor market. Wages differ substantially across workers with similar characteristics and similar jobs, and turnover is lower in industries that pay higher wages (Krueger and Summers 1988). Larger firms tend to pay higher wages (Brown and Medoff 1989). Low-wage firms appear to have a substantial number of vacancies; for example, more than 40 percent of surveyed restaurant operators reported they had at least one position vacant for more than two weeks in December 1989 (National Restaurant Association 1990). Vacancies should not occur in perfectly competitive labor markets because wages adjust to equilibrate labor demand and supply. In the monopsony framework, firms with vacancies do not offer a higher wage because they would also have to pay the higher wage to current employees. A minimum wage increase that forces firms to raise wages may attract more workers to fill vacancies, causing employment to rise.

Dynamic monopsony models may explain some of the recent research findings. However, additional evidence is necessary. Empirical work has not tested whether hires increase and vacancies and quits fall as the minimum wage increases. Establishment-level data on vacancies, hires, and quits as well as wages and employment before and after a minimum wage hike need to be collected in order to test the dynamic monopsony model. Further research on prices is also needed if dynamic monopsony models are invoked to explain positive employment effects of minimum wage increases. Like the traditional monopsony model, these models imply that output increases when employment increases, and there is little evidence of the expected accompanying fall in prices.

Endogeneity. All of the above models simply assume that a wage floor is imposed or raised, ignoring how the minimum wage is determined. However, the level of the minimum wage may depend on the expected effect on employment. Federal and state minimum wages in the United States are primarily set by politicians, who are likely to be concerned about the potential negative effects of a minimum wage increase and its effects on their reelection chances. Politicians may opt not to raise the minimum wage if the disemployment effect will be large or if it will substantially erode businesses' profits. If any negative effects are likely to be minimal, perhaps because the economy is doing well and average wages are rising, politicians may raise the minimum wage in an effort to appeal to low-wage voters without losing too much support from businesses.

Lawmakers' statements suggest that the timing of minimum wage increases depends on economic conditions and on the likely impact. For example, the governor of Connecticut, approving an $0.88 increase in the state's minimum wage in 1987, observed that "the state with the best economy in the nation can afford this minimum wage."[9] The commission that determines state minimum wages in California refused to increase the state's wage floor in 1993 because "any increase would further damage the state's ailing business climate."[10] The New Jersey state legislature and governor approved in 1990 a $0.90 minimum wage increase to take place in 1992; when the state's economy slipped into a recession during the intervening period, the legislature tried to delay part of the scheduled increase.[11]

If minimum wage increases are designed to occur when they will have minimal impact, it is not surprising that researchers have had difficulty finding negative employment effects. If the level of the minimum wage depends on its effect on employment, the minimum wage is endogenous with respect to employment. Traditional

techniques that measure the effect of the minimum wage on employment that do not account for this endogeneity will yield incorrect results. As discussed earlier, the typical time-series model regresses the teen employment rate on a measure of the minimum wage and other variables. This estimation technique requires that the minimum wage variable be uncorrelated with the error term in the regression, or that shocks to teen employment do not affect the minimum wage. But if minimum wage increases occur when teen employment is high, the minimum wage is likely to be positively correlated with the error term. The estimated effect of the minimum wage will then be positively biased. Failure to control for endogeneity bias causes the disemployment effect to be underestimated. Similarly, difference-in-differences comparisons between employment changes in states that raise their minimum wage and states that do not may be biased if teen employment is initially growing faster in the states that raise their minimum wage and this growth prompts the hike. These cross-state comparisons are particularly susceptible to endogeneity bias since there must be some differences between the states that lead some but not others to raise their minimum wage.

One method of controlling for endogeneity bias is finding a source of variation in minimum wages that is unrelated to economic conditions. More formally, at least one variable that is strongly correlated with the minimum wage but uncorrelated with the error term in the employment regression is needed to identify the true effect of the minimum wage on employment. This econometric technique is termed instrumental variables estimation. In effect, an instrumental variable (which is uncorrelated with the residual in the employment regression) is substituted for the endogenous variable. Neumark and Wascher (1992) attempt to identify the effect of minimum wage increases by using the average minimum wage in neighboring states as an instrument for state minimum wage levels. They obtain more negative estimates of the effect of the minimum wage on employment among teens and young adults, a finding that is consistent with endogeneity bias, but the estimates are imprecise. In addition, the average minimum wage of neighboring states may be a problematic instrument if contiguous states move together in business cycles. Further research using more powerful instruments that are uncorrelated with economic conditions is necessary to support the claim that endogeneity bias causes nonnegative estimates of the employment effect of the minimum wage.

An alternative method of controlling for endogeneity bias is to find a minimum wage increase that is plausibly unrelated to economic conditions. Card (1992a) offers an example of such research. Cross-state comparison of the effects of the 1990 federal minimum wage hike should be immune from endogeneity bias since the minimum wage increase was imposed on the states by the federal government. In addition, the states most affected by the federal minimum wage increases were those that had opted to keep their state minimum wages at low levels, and the states that were least affected had already increased their state minimum wages above the federal level; the endogeneity hypothesis implies that these high-wage states were growing faster than low-wage states before the minimum wage hike. Card finds no evidence of negative employment effects in low-wage states relative to high-wage states; this lack of evidence does not support the possibility that endogeneity underlies the failure to find negative effects in other studies.

Recent findings that minimum wage increases do not appear to affect employment adversely should be taken as the starting point for a larger examination of the effects of the minimum wage level.

Additional Theories. Firms could attempt to offset an increase in their wage bill due to a minimum wage hike by reducing other labor costs, such as fringe benefits and training. If firms can completely offset a minimum wage increase by cutting other costs, they might not reduce employment. Although this theory is plausible, it is unlikely that low-wage employers have been able to substantially reduce nonwage labor costs because they provide relatively little in fringe benefits and training. Alpert (1986) notes that restaurant workers, for example, received about 20 percent of the fringe benefits income received by other workers. Alpert finds that restaurants slightly reduced fringe benefits when the minimum wage rose during the 1970s, but Katz and Krueger (1992) find no evidence that fast-food restaurants that were more affected by minimum wage increases cut fringe benefits relative to higher-wage establishments. The prevalence of on-the-job training also appears to be low in low-wage, entry-level jobs; Lynch (1992) reports that only 4.2 percent

8. Other papers include Lang (1994), who develops a bilateral search model in which a minimum wage can raise total employment but cause employment among low-wage workers to fall, and Rebitzer and Taylor (1995), who present an efficiency wage model in the context of a minimum wage that has implications similar to a monopsony model.

9. Bureau of National Affairs, Daily Labor Report, *August 17, 1987, A2.*

10. Ibid., August 31, 1993, A9.

11. Ibid., March 25, 1992, A3.

of young adults who did not complete college reported receiving on-the-job training.

Another potential reason why employment might not fall when the minimum wage increases is that employers might not comply with the new law. For example, Card and Krueger (1995) find that the noncompliance rate, or the percentage of workers earning less than the minimum wage who should be paid at least the wage floor, rose from 31 percent to 46 percent in California after the state's 1988 minimum wage hike. However, of the studies finding that minimum wage increases did not reduce employment that also investigate the effect on wages, all find that the increases had a substantial effect on the distribution of wages, indicating that most employers complied with the law.

Some employers do not have to comply with minimum wage laws, and economists have posited that displaced workers might move to these firms when the minimum wage increases. The federal minimum wage law does not apply to some very small firms and to some agricultural establishments. Workers who are laid off from jobs covered by the minimum wage might find jobs in the uncovered sector, explaining why total employment appears unaffected. However, the uncovered sector composes only a small percentage of employment, making it unlikely that it could absorb many displaced workers. This theory also cannot account for the finding that employment at specific restaurants does not fall when the minimum wage increases.

Conclusion

Recent research has challenged the conventional wisdom among economists that increases in minimum wages lower employment among low-wage workers. Previous research generally found small but significant negative effects of higher minimum wages on low-wage workers, particularly teenagers. Although some studies, such as Deere, Murphy, and Welch (1995), continue to find negative effects, the findings of Card, Katz, and Krueger and others raise the question of whether the minimum wage can be raised moderately without reducing employment. If these findings are correct, economists may need to reconsider their views of how labor markets work.

Two of the models explored here seem unlikely to explain why minimum wage increases do not reduce employment: price effects and traditional monopsony. Price increases can ameliorate the disemployment effects of a minimum wage hike in the competitive model, and some studies find that prices in the restaurant industry rise when the minimum wage increases. However, demand must be completely inelastic—and it is not likely to be—for price changes to completely offset the negative employment effects. Most economists are unwilling to accept the monopsony model because they believe that few low-wage employers are large enough to face an upward-sloping labor supply curve.

Several models can explain why employment does not appear to fall when the minimum wage increases, but further research is needed to determine their validity. The substitution model, which posits that employers replace lower-skilled workers with higher-skilled workers when the minimum wage increases, can predict that total employment is unchanged. Some studies find results consistent with this model, but research with establishment-level data on whether employers substitute among workers with different skill levels and leave total employment unchanged when the minimum wage increases is needed. Dynamic monopsony models with frictions in search processes can predict that employment increases, decreases, or is unaffected by a minimum wage hike. These models are appealing because they can explain many stylized facts about the labor market. However, their ambiguous predictions of the effect of a minimum wage increase make them difficult to test. Research that examines the effect of minimum wage increases on vacancies, quits, and hires is needed. The endogeneity hypothesis that minimum wage increases occur when disemployment effects are minimal also requires further research to establish its validity.

Employment is not the only area worthy of further research on the effects of minimum wage increases. The distributional consequences of minimum wage increases are at least as important as the employment effects, particularly if higher-skilled workers displace lower-skilled workers when the minimum wage rises.[12] Minimum wage increases may also slow the rate of small business formation, a possibility that has not received much attention in the economics literature. The recent findings that minimum wage increases do not appear to affect employment adversely should be taken as the starting point for a larger examination of the effects of the minimum wage level rather than an end to the debate.

12. For research on the distributional effects of minimum wage increases, see Burkhauser, Couch, and Wittenburg (1996), Addison and Blackburn (1997), and Neumark and Wascher (1997) .

REFERENCES

Aaronson, Daniel. 1997. "Price Pass-Through and Minimum Wages." Federal Reserve Bank of Chicago Working Paper No. 97-3, July.

Addison, John T., and McKinley L. Blackburn. 1997. "Minimum Wages and Poverty." University of South Carolina. Photocopy.

Alpert, William T. 1986. *The Minimum Wage in the Restaurant Industry.* New York: Praeger.

Brown, Charles. 1995. "Comment." *Industrial and Labor Relations Review* 48 (July): 828–30.

Brown, Charles, Curtis Gilroy, and Andrew Kohen. 1982. "The Effect of the Minimum Wage on Employment and Unemployment." *Journal of Economic Literature* 20 (June): 487–528.

Brown, Charles, and James Medoff. 1989. "The Employer Size-Wage Effect." *Journal of Political Economy* 97 (October): 1027–59.

Brown, Douglas M. 1990. "The Restaurant and Fast Food Race: Who's Winning?" *Southern Economic Journal* 56 (April): 984–95.

Burdett, Kenneth, and Dale T. Mortensen. 1989. "Equilibrium Wage Differentials and Employer Size." Northwestern University Working Paper No. 860, October.

Burkhauser, Richard V., Kenneth A. Couch, and David C. Wittenburg. 1996. "'Who Gets What' from Minimum Wage Hikes: A Re-Estimation of Card and Krueger's Distributional Analysis in *Myth and Measurement: The New Economics of the Minimum Wage." Industrial and Labor Relations Review* 49 (April): 547–52.

Card, David. 1992a. "Using Regional Variation in Wages to Measure the Effects of the Federal Minimum Wage." *Industrial and Labor Relations Review* 46 (October): 22–37.

———. 1992b. "Do Minimum Wages Reduce Employment? A Case Study of California, 1987–89." *Industrial and Labor Relations Review* 46 (October): 38–54.

Card, David, Lawrence F. Katz, and Alan B. Krueger. 1994. "Comment on David Neumark and William Wascher, 'Employment Effects of Minimum and Subminimum Wages: Panel Data on State Minimum Wage Laws.'" *Industrial and Labor Relations Review* 47 (April): 487–97.

Card, David, and Alan B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84 (September): 772–84.

———. 1995. *Myth and Measurement: The New Economics of the Minimum Wage.* Princeton, N.J.: Princeton University Press.

———. 1998. "A Reanalysis of the Effect of the New Jersey Minimum Wage Increase on the Fast-Food Industry with Representative Payroll Data." Princeton University, Industrial Relations Section, Working Paper No. 393.

Currie, Janet, and Bruce C. Fallick. 1996. "The Minimum Wage and the Employment of Youth: Evidence from the NLSY." *Journal of Human Resources* 31 (Spring): 404–28.

Deere, Donald, Kevin M. Murphy, and Finis Welch. 1995. "Employment and the 1990–1991 Minimum Wage Hike." *American Economic Review Papers and Proceedings* 85 (May): 232–37.

Frey, Bruno S., Werner W. Pommerehne, Friedrich Schneider, and Guy Gilbert. 1984. "Consensus and Dissension among Economists: An Empirical Inquiry." *American Economic Review* 74 (December): 986–94.

Fuchs, Victor R., Alan B. Krueger, and James M. Poterba. 1997. "Why Do Economists Disagree about Policy? The Role of Beliefs about Parameters and Values." National Bureau of Economic Research Working Paper No. 6151, August.

Hamermesh, Daniel S. 1995. "Comment." *Industrial and Labor Relations Review* 48 (July): 835–38.

Katz, Lawrence F., and Alan B. Krueger. 1992. "The Effect of the Minimum Wage on the Fast-Food Industry." *Industrial and Labor Relations Review* 46 (October): 6–21.

Kennan, John. 1995. "The Elusive Effects of Minimum Wages." *Journal of Economic Literature* 33 (December): 1949–65.

Kim, Taeil, and Lowell J. Taylor. 1995. "The Employment Effect in Retail Trade of California's 1988 Minimum Wage Increase." *Journal of Business and Economic Statistics* 13 (April): 175–82.

Krueger, Alan B., and Lawrence H. Summers. 1988. "Efficiency Wages and the Inter-Industry Wage Structure." *Econometrica* 56 (March): 259–93.

Lang, Kevin. 1994. "The Effect of Minimum Wage Laws on the Distribution of Employment: Theory and Evidence." Boston University. Photocopy.

Lynch, Lisa M. 1992. "Private-Sector Training and the Earnings of Young Workers." *American Economic Review* 82 (March): 299–312.

Machin, Stephen, and Alan Manning. 1992. "The Effects of Minimum Wages on Wage Dispersion and Employment: Evidence from the U.K. Wages Councils." *Industrial and Labor Relations Review* 47 (January): 319–29.

National Restaurant Association. 1990. *Foodservice Employers and the Labor Market.* Washington, D.C.: National Restaurant Association.

Neumark, David, and William Wascher. 1992. "Employment Effects of Minimum and Subminimum Wages: Panel Data on State Minimum Wage Laws." *Industrial and Labor Relations Review* 46 (October): 55–81.

———. 1995a. "Minimum Wage Effects on Employment and School Enrollment." *Journal of Business and Economic Statistics* 13 (April): 199–206.

———. 1995b. "The Effects of Minimum Wages on Teenage Employment and Enrollment: Evidence from Matched CPS Surveys." National Bureau of Economic Research Working Paper No. 5092, April.

———. 1997. "Do Minimum Wages Fight Poverty?" National Bureau of Economic Research Working Paper No. 6127, August.

PRESIDENT. 1996. Statement by the President, Washington, D.C., July 9. Available on-line at http://library.whitehouse.gov.

REBITZER, JAMES B., AND LOWELL J. TAYLOR. 1995. "The Consequences of Minimum Wage Laws: Some New Theoretical Ideas." *Journal of Public Economics* 56 (February): 245–55.

WELCH, FINIS. 1995. "Comment." *Industrial and Labor Relations Review* 48 (July): 842–49.

WELLINGTON, ALISON. 1991. "Effects of the Minimum Wage on the Employment Status of Youths: An Update." *Journal of Human Resources* 26 (Winter): 27–46.

WESSELS, WALTER J. 1997. "Minimum Wages and Tipped Servers." *Economic Inquiry* 35 (April): 334–49.

New Drilling Technology

THOMAS J. CUNNINGHAM AND WHITNEY MANCUSO
Cunningham is vice president in charge of the regional section of the Atlanta Fed's research department. Mancuso is an analyst in the regional section. *They thank Jimmy Stone of Stone Energy, David Guidry of Guico Machine Works, and Frank King and Will Roberds for insightful discussion and comments.*

MACROECONOMIC POLICYMAKERS FACE TWO FUNDAMENTAL WORRIES. THE FIRST IS WHETHER THE ECONOMY IS ON A GOOD OR DESIRABLE GROWTH PATH. ECONOMIC FORECASTS USUALLY ANSWER THIS QUESTION, SHOWING THE PROBABLE COURSE OF THE ECONOMY OVER THE NEXT FEW QUARTERS IF SOME BASIC SET OF CONDITIONS REMAINS STABLE OR AT LEAST BEHAVES AS PREDICTED. THE SECOND FUNDAMENTAL WORRY IS WHETHER THESE BASIC ASSUMED CONDITIONS MIGHT CHANGE RADICALLY AND QUICKLY, PUSHING THE ECONOMY AWAY FROM ITS FORECAST PATH. THESE UNFORECASTABLE DISTURBANCES ARE KNOWN AS ECONOMIC SHOCKS.

Shocks are mostly of concern when the potential outcome is bad. The most infamous example of a negative economic shock occurred in the 1970s, when the Organization of Petroleum Exporting Countries (OPEC) imposed oil embargoes. Serious negative economic consequences resulted and proved especially problematic for policymakers because of their sudden, essentially unforeseen onset.

Shocks do not, however, have to be negative. Indeed, this article suggests that at the moment the energy extraction industry is in the midst of a very positive shock caused by a combination of two new technologies. These technologies emerged very quickly, and while either technology by itself would have improved the drilling process, as discussed below, it is the combination of the innovations that has served to substantially reduce energy extraction costs in certain types of geological formations.

The oil and gas industry has seen a series of dramatic technical developments in two distinct areas: three-dimensional imaging and directional (or horizontal) drilling. Combined, these two revolutionary techniques have significantly lowered the net extraction costs of oil and made feasible the reopening of wells and fields that had ended their economic usefulness under the old technology. The result has been a surge in the energy sector that is all the more significant because it is being pushed not by a spike in prices but rather by a drop in production costs and may therefore be lasting.

So far this positive technology shock has most affected the oil and gas industry in the Gulf of Mexico. The new technologies are particularly advantageous for the kinds of fields characteristic of the gulf, and their early application there can yield information about what to expect as use of the technologies spreads. The impacts on extraction and exploration costs in other fields in the United States and abroad are likely to be significant.

The purpose of this article is to examine these changes in the energy extraction industry, focusing on Louisiana for examples of the particular benefits in the Gulf of Mexico. This regional focus is appropriate for two

reasons. First, the geology of energy deposits in the gulf is relatively complex, so the benefits of the new technology are even greater in its fields than in other parts of the world where the formations are simpler. Second, the drilling technologies have their greatest payoff in offshore fields like Louisiana's, where the costs of exploratory drilling are greatest.

The article is of direct interest at the regional level in terms of what the new technologies promise for Louisiana's economy. In a larger sense, this inquiry offers the rare opportunity to observe a major technological revolution taking place in a mature industry. The oil and gas extraction business, an important and already well capitalized industry, is experiencing a major, identifiable, positive technology shock.

The discussion begins with a short history of drilling in the Gulf of Mexico, a microcosm of the industry's development that helps set the stage for discussing the significance of the new technologies. A close look at the two innovations in drilling technology and what they mean for the energy industry follows. The discussion concludes with a return to Louisiana, why these developments matter so directly to the state, and what Louisiana's experience might, in turn, imply for other oil-producing regions.

A History of Offshore Drilling in Louisiana

Offshore oil and gas exploration in the Gulf of Mexico began off the Louisiana coast. By the early 1930s the major oil companies dominated onshore production and prospects in Louisiana, and offshore drilling remained too risky an experiment for them to undertake. Instead, small independent oil companies took up the challenge and built the first wooden oil platform in waters near Creole, Louisiana, in 1933 ("1947" 1997). Eventually, wooden platforms gave way to cheaper and sturdier steel platforms, allowing safer work farther from shore. In 1947 one of the small independents, Kerr McGee, hit oil nine miles off Louisiana's outer islands ("Milestones" 1997). Constrained by limited finances, Kerr McGee built a small platform in the deep water and tendered a U.S. Navy yard freighter to it for support. The ship had space onboard for crews to sleep, and drilling operations could continue for longer periods each day. The success of Kerr McGee's operation encouraged drilling activity in the gulf, but costs remained high.

It was not until the end of World War II that naval architects were able to turn their attention to the oil industry. The first technical push was for simple well-to-well mobility and the ability to work in greater water depths. Eventually, the industry moved from fixed drilling platforms such as jackups and submersibles, which rest on the ocean floor, to semisubmersibles and drillships, which are held in place by a set of anchors and thus capable of working in much deeper water. These innovations allowed operations to move easily when necessary, further reducing costs. During the 1950s the settlement of a long-running tideland ownership dispute between the federal government and states, which facilitated offshore leasing by establishing jurisdiction, also boosted drilling activity, as did the successful introduction of new marine seismic techniques ("1947" 1997).

Revolutionary techniques have significantly lowered the net extraction costs of oil and made feasible the reopening of wells and fields that had ended their economic usefulness under the old technology.

During the mid-1950s the offshore industry in the gulf began to slow down. Nuclear energy was emerging, and oil was plentiful worldwide. Crude oil prices, in relative terms, dropped dramatically, and domestic price controls provided additional distortions to the domestic production market (Bennett, Cole, and Dym 1980–81). At the same time, more and more operators in the gulf were hitting dry holes. The resulting slowdown lasted until 1970, when the excess oil supply had been extinguished and a market had been established for natural gas, which is relatively abundant in the gulf.

The first wide-scale oil embargo among OPEC nations took place in 1973, and supply problems increased as lines formed at gas stations and pump prices rose dramatically. Domestic price controls began to be lifted, and drilling activity in the gulf picked up, spurred further by rising prices in the late 1970s. (Chart 1 shows the path of oil prices and related employment in Louisiana beginning in 1975.) At the same time, introduction of a seismic technology that allowed a look essentially straight down into the earth made it much easier to find likely concentrations of potential hydrocarbon deposits and for the first time enabled searches for oil beyond the continental shelf.[1]

The second OPEC oil embargo resulted from a revolution and political turmoil in Iran during 1978 and 1979 that significantly cut the country's oil production. The embargo continued as Iraq invaded Iran in 1980. Oil prices rose sharply, and drilling activity in the Gulf of Mexico followed suit; however, during this time offshore

1. This seismic technology allowed a detailed analysis of a change in the characteristics of the reflected waves, focusing on subsurface areas where the geology was most consistent with hydrocarbon deposits.

CHART 1 Louisiana Oil Prices and Related Employment, 1975–96

Source: Price of Louisiana oil from Louisiana Department of Natural Resources (1996, 21); mining employment statistics from the U.S. Bureau of Labor Statistics

costs were also rising, in part because of increased environmental concerns regarding offshore oil development and because new discoveries were getting smaller. The industry expected that high oil prices would remain in place for some time, but in 1985 Saudi Arabia decided it would no longer restrict its supply of oil. Instead, Saudi Arabia decided to adopt a high production rate, which meant immediate increases in crude oil supply and a sharp decline in prices ("1947" 1997). When oil prices plunged in the fall of 1985, operators in the Gulf of Mexico could not continue to produce oil at those prices because of the high cost of extraction and its associated maintenance offshore. Hundreds of leases in the Gulf of Mexico were returned to the government, drilling was cut back significantly, and employment in the sector declined.

By 1988 oil prices began to recover somewhat, but major oil companies were no longer interested in the deep water of the gulf. However, independent (and largely local) operators remained interested and went heavily in debt purchasing large tracts of offshore extraction rights from the major companies. In an effort to make their purchases pay off quickly, exploration drilling (in search of new deposits) gave way to development drilling to more rapidly exploit known deposits, and independent operators ultimately drilled more wells in the gulf than the major oil companies did. Somehow, the independents had to cut drilling costs significantly, and they did so by erecting low-cost minimal platforms, as described earlier, and by refurbishing cheap old platforms. Despite efforts at cost-cutting and switching from drilling for oil to drilling for natural gas, 1992 and 1993 were painful years for gulf operators as gas prices fell dramatically and oil prices dropped to below $18 per barrel. Finally, however, in 1994 the merging of several new technologies, in tandem with rising oil and gas prices, culminated in both lower drilling costs and higher profits for gulf operators.[2] It is to these new technologies that the discussion now turns.

The Symbiosis in Two New Technologies

Around 1994 two relatively new groups of technologies were combined to drastically reduce oil and gas extraction costs. The first of these is three-dimensional seismic imaging, and the second is controlled directional (steerable, as opposed to the conventional vertical) drilling.[3]

Three-dimensional seismic imaging is a combination of recent innovations that provides geophysicists with very large quantities of seismic data for greater precision in defining and creating images of possible deposits of oil and gas in deep geologic formations. Scientists can model and identify those formations likely to contain extractable hydrocarbon deposits, hence reducing uncertainty—and thus costs—in the exploration process.

Directional drilling proceeds with fair precision along a long and complex path. This innovation offers two related but distinct advantages over a traditional vertical well. First, it allows access and more precise exploitation of complex deposits identified through three-dimensional imaging. Being able to accurately identify deep formations would be of little use if drilling technology did not permit access to them. On the other hand, of course, the ability to precisely locate a drilling path would be of limited use if deposits cannot be precisely identified.

The second advantage to steerable drilling techniques is directly related to drilling offshore. Because the drill can now be steered horizontally (or in any direction),

the ultimate location of the oil deposits no longer has to be particularly close to the drilling platform. The oil- or gas-bearing formations in the Gulf of Mexico are typically quite complex and not economically accessible by conventional wells. Now that drilling is no longer limited by the position of the rig, one offshore drilling platform can effectively exploit many different deposits in a wide subsea area. As a result, offshore drilling costs less.

Three-Dimensional Imaging. Three-dimensional seismic imaging combines two important innovations: relatively inexpensive computing power and some geophysical algorithms that can interpret seismic data to form 3-D images. The two technologies have depended on each other to make 3-D imaging practically—or economically—feasible (Neff and Thrasher 1993).

The geophysical problem in seismic imaging is that different geologic formations allow seismic pulses to travel at different rates of speed. These differences in rates of wave propagation and reflection are what allow the images to be made in the first place, much like radar depends on the reflective properties of different materials to produce an image. However, as the geologic formations become more complex, either because the imaging process is probing more deeply or because it is looking at a particularly complex formation, it becomes exponentially more difficult to extract a true image. In addition, some geologic formations distort sonic waves so much that traditional two-dimensional imaging techniques are overwhelmed by the distortion. Salt deposits, a common feature in the Gulf of Mexico, are such a formation (Neff and Thrasher 1993). Addressing these issues is one of the more powerful applications of 3-D imaging in the gulf.

The 3-D imaging process itself is dependent upon several key recent technical innovations. Instead of surveying a specific surface block and creating a two-dimensional image of what is beneath (as occurs in surveying for conventional well-shaped deposits), the three-dimensional process requires information from all the blocks around the block of interest. Moreover, the finer the ultimate precision of the image is, the more raw data is needed for interpretation. Only in the last decade has adequate computing and data power made it feasible to economically gather, store, and manipulate such massive quantities of data.

Just as important in offshore work is that the process of gathering the survey data depends on knowing the precise location of data sensors. Low-cost and highly accurate geopositioning satellite technology, available only within the last decade, provides an enormous improvement in accuracy and efficiency over the days when one ship carried one sensor by allowing a single ship to tow a lengthy array of sensors, the location of which can be known with some certainty.

Transforming the raw data into an image requires an implementable mathematical formula—an algorithm. Researchers have recently developed improved algorithms for interpreting seismic data that can yield detailed three-dimensional images. In addition, the availability of supercomputers small and robust enough to be taken into the field has made the 3-D seismic mapping process economically viable. The new technology allows a resolution fine enough for 3-D mapping of deep geologic formations and therefore suitable for the guidance of exploratory and developmental wells. This process is continuously being refined and with enough success that the current leading-edge imaging technology is referred to as 4-D.

The oil- or gas-bearing formations in the Gulf of Mexico are typically quite complex and not economically accessible by conventional wells.

On an operational level, the new technologies are not wholly without drawbacks. Because both technologies are relatively sophisticated, there is a high premium on specialized human capital. During the last drilling slowdown, many experienced workers left the industry, and their defection has placed an even greater premium on specific skills in the industry. Also, the technology offers its greatest comparative advantage in deep-water exploration, and drilling off the continental shelf requires large investments and relatively long planning lags.

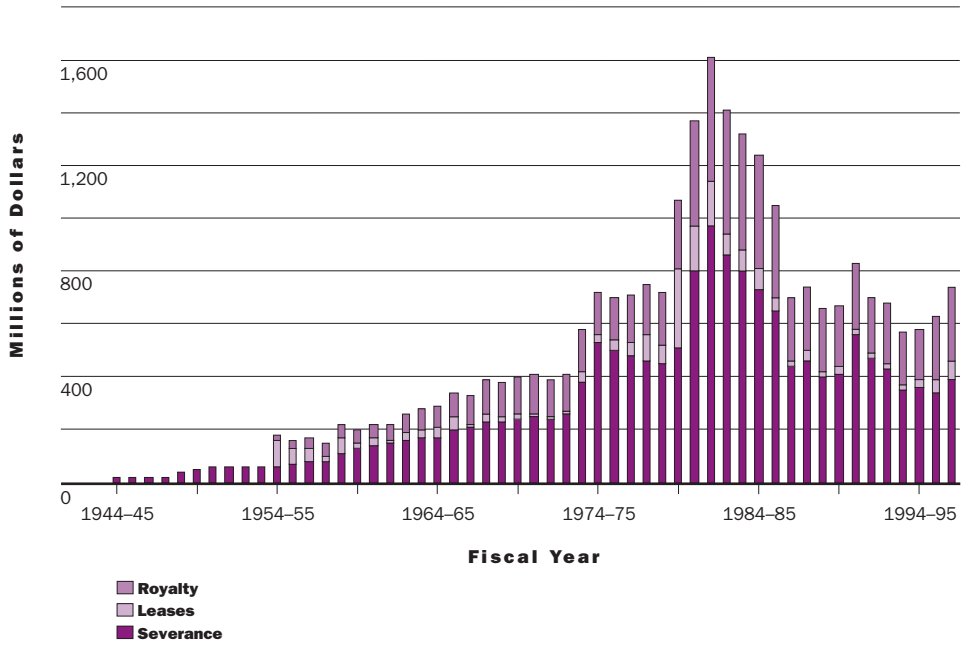
Directional Drilling and the Problem of Irregularly Shaped Deposits. Drilling used to be, literally, a fairly straightforward matter. Although not necessarily strictly vertical, wells were largely straight-line constructions. This approach works well if the oil or gas deposit happens to be held in a conveniently shaped straight formation or is so vast that the precise placement of the well does not really matter, as in the Persian Gulf and some of the early Texas oil fields, for example. In the Gulf of Mexico, however, the irregular shapes of formations necessitated a great deal of drilling to get at the deposits, hence making it economically infeasible to extract most of the deposit.

The limitation of previous technology was that it provided no way to effectively steer the bit once it was

2. See the Web site of the Center for Energy Studies, Louisiana State University (www.enrg.lsu.edu), for more energy statistics and additional oil industry links.

3. For a more detailed description of these technologies, see Oil & Gas Journal *on-line at www.ogjonline.com.*

CHART 2 Louisiana Revenue from Oil and Gas Production, 1944–97



Source: Louisiana Department of Natural Resources unpublished data

far below ground. One issue was how to determine exactly where the bit was and where it was going and communicate this data back to the driller. Unrefined versions of the two basic systems for navigating a drill—magnetic and gyroscopic—were conceptually available earlier than the recent boom, but both methods had fairly complex telemetry requirements. This problem was overcome for only the shallowest of wells until, not entirely coincidentally, about the time the laptop computer appeared. Previously, drillers monitored geologic conditions by examining the drilling mud, a method both imprecise and not timely.

The results provided by magnetic drill navigation, which functions much like a complex compass, can be distorted by local geologic formations, and these change as drilling proceeds. Solving this problem called for a model of different formations' magnetic distortion and an accurate picture of the formations the drill is passing through so that corrections for distortions could be made. Until the advent of 3-D imaging provided a usefully complete picture of formations that would alter the magnetic locational data, this requirement was not feasibly met.

The second method of navigating a drill is to use a gyroscopic guidance system analogous to the inertial guidance systems in airplanes, in which analysis of momentum on a gyroscope can provide relative location information without any external signals or information. Because gyroscopic guidance is self-contained, it can avoid the magnetic distortion problem, but the technology is inherently more complex and difficult to make reliable. A continuous, real-time gyroscopic guidance system that was rugged, reliable, and accurate enough to greatly enhance the precision of the steerable drill became commercially available in 1995.

These developments overcame the engineering problem of having drills make fairly precise and sharp turns at arbitrary depths. With their availability, the entire process of steering the drilling of a well to fit whatever line was required became feasible.

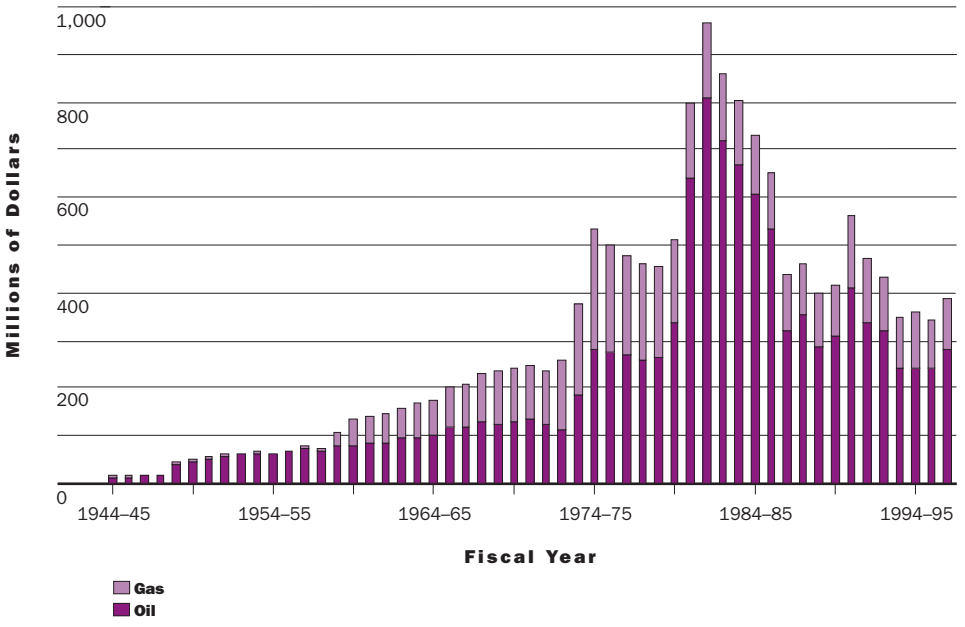
Implications for Louisiana

The application of these technical innovations in the Gulf of Mexico has resulted in something of a mini-boom. A broader application may cause positive, though probably lesser, shocks in other oil-producing regions and could eventually alter the world's energy economy. To get some idea about how economies and governments' fiscal condition in other oil-producing regions might react, an analysis of effects on the Louisiana economy is in order.

The energy extraction industry in Louisiana has expanded rapidly over the last few years. This growth has occurred despite considerable volatility in the price of a barrel of crude—a mild run-up in 1995 and 1996 and, more recently, a run back down, with little net price movement from the beginning of the gains in drilling in 1995 (see Chart 1).

Louisiana was a prime beneficiary of the last surge in energy extraction activity, during the high oil prices of

CHART 3 Louisiana Severance Tax Revenue from Oil and Gas, 1944–97



Source: Louisiana Department of Revenue and Taxation unpublished data

the 1970s and early 1980s. The energy industry directly provided a large portion of the state's overall tax revenue as well as considerable personal and business income to the workers and firms employed in, and in support of, the energy extraction process. Given the sudden increase in the real cost of energy in 1974, this result is not especially surprising. In turn, when the real (and nominal) price of oil began to fall in the mid-1980s, Louisiana fell on comparatively hard times as energy income diminished.

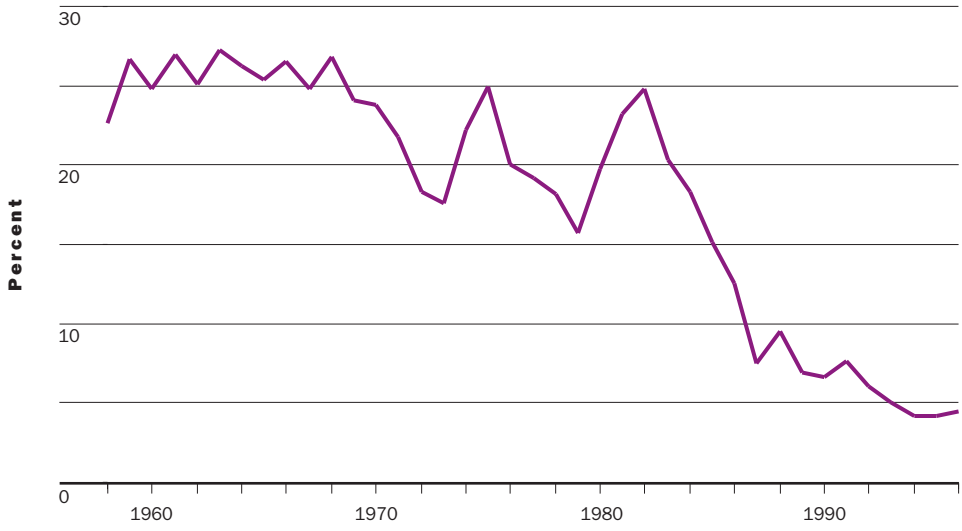
Over the past year, although oil prices have fallen, Louisiana's energy industry has continued its prosperity begun in the mid-1990s. This time around, however, the energy industry's continued strength has not been powered by some externally imposed increase in the price of oil or gas, as discussed earlier. Growth in Louisiana's mining sector—essentially all oil and gas extraction—picked up dramatically in late 1995, and that rate of growth has only incrementally slowed since then. At the same time, the price of oil rose mildly through mid-1997 but then fell off rapidly, and it did so with no concomitant spillover into mining employment growth.

Good times in the oil and gas industry have historically meant good times for Louisiana. Will the current positive energy shock based on technological innovation have implications for the state similar to those of an increase in the price of energy? To date at least, the energy sector of the state's economy and state tax revenues has clearly benefited.

Louisiana had traditionally relied heavily upon the oil and gas industry for revenue; however, that reliance waned over the last decade in response to the industry's slump. The state collects tax revenue directly from the industry in the form of severance taxes, royalty payments, and lease fees. Indirectly, the state collects moneys from corporate and personal income taxes, corporate franchise taxes, sales taxes, and property taxes, all of which may be influenced by the condition of the local oil and gas industry. The postwar history of oil and gas total dollar revenue collections for Louisiana is shown in Chart 2, broken down by tax classification. Chart 3 shows the composition of severance tax collections in Louisiana, divided between oil and gas.

During the 1960s, Louisiana received nearly 25 percent of its tax revenue funds from oil and gas severance taxes, royalty payments, and lease fees (see Chart 4). During the 1970s the proportion of revenues from the oil and gas industry was slightly more volatile (as were product prices) but remained strong (as, again, did the product price). When oil prices began to tumble in the 1980s due to the increase in world oil supply, Louisiana's state oil and gas revenues began to drop largely due to the falloff in crude oil prices. The state lost additional income as drilling firms and their suppliers closed up shop and laid off employees. By 1990 oil and gas revenue accounted for slightly less than 7 percent of Louisiana's total state revenues. Despite a minor resurgence in oil and gas exploration and production in the Gulf of Mexico during the first half of the 1990s, the petroleum industry's share of total state revenue has continued to dwindle. However, the total amount of revenue con-

CHART 4
Louisiana Oil and Natural Gas Severance Taxes, Royalty Payments, and Lease Fees as a Percentage of Total State Revenue, 1958–96



Source: Louisiana Department of Natural Resources unpublished data and U.S. Bureau of the Census, *Government Finances,* various issues

tributed to state coffers by oil and gas extraction has remained relatively stable. The decline in share may be due to substantial growth in the total overall amount of revenue that Louisiana brings in.

Even though the petroleum industry has once again become interested in drilling prospects in the waters off Louisiana's coast, oil and gas production remain far below that of the late 1970s, and relatively low oil prices have prevented the state from realizing a windfall in tax receipts. Additionally, since the late 1970s a greater proportion of oil and gas has been extracted from deeper waters that are the jurisdiction of the federal government and are thus not directly subject to state taxes.[4] According to the state department of natural resources, Louisiana has largely been unsuccessful at directly recovering what it believes are costs of creating infrastructure to support the now more intense offshore operations. Several attempts have been made at taxing oil and gas retrieved from federal waters and imported into the state for refinement. These efforts have been limited, however, largely because of the fear that over the long run taxation would divert refining investment to other Gulf Coast states (Pulsipher 1990).

In addressing the relative position of Louisiana in terms of energy industry taxation, researchers at Louisiana State University in 1993 found that the tax burden on firms that find and produce oil and natural gas in Louisiana did not differ significantly overall from that in competing states (Pulsipher, Baumann, and Iledare 1993). Louisiana does, however, limit its revenues. The state taxes oil at a higher rate than natural gas, although for mostly geophysical reasons Louisiana produces roughly twice as much gas as oil.

Severance tax collections make up the largest portion of direct revenue collected from the oil and gas industry. A severance tax is generally levied on all natural resources extracted from the soil or water and in Louisiana is paid by the natural resource owner. More than 95 percent of the state's severance tax collections are attributed to oil and gas extraction. In 1910 the state administered its first severance tax on oil and gas production through a minimal occupational license tax, and then in 1922 constitutional authority was given for a severance tax (Louisiana Department of Revenue and Taxation 1997, 151). Since that time there have been many changes to the tax rate, the most significant being that collections have moved from a volume basis to a percentage-of-value basis.

Oil production accounts for the majority of severance tax collections, and the amount of revenue the state collects from oil is more dependent upon the price of oil than on the amount of oil produced. Since 1974 most oil wells have been taxed at the rate of 12.5 percent of value produced (Louisiana Department of Revenue and Taxation 1997, 151). According to the Louisiana Department of Natural Resources, "At constant production, the State Treasury gains or loses about $20 million of direct revenue from oil severance taxes and royalty payments for every $1 per barrel change in oil prices. This figure rises to $30 to $40 million per dollar change when indirect revenue impacts are included" (Louisiana Department of Natural Resources 1997). Severance tax

collections from oil peaked during the 1981–82 fiscal year at more than $809 million. Similarly, the average wellhead price of Louisiana crude oil rose to its highest level at $33.84 per barrel during the same fiscal year, which runs from July to June. Oil production at that time was rather modest when compared with its highs of the mid-1970s. At that time the oil severance tax was collected at the rate of $0.18 to $0.26 per barrel based upon its weight. In 1986, when the bottom dropped out of oil prices, the average Louisiana wellhead price for crude oil during the fiscal year fell to $15.43 per barrel. Likewise, collections from the oil severance tax fell to slightly more than $318 million during the 1986–87 fiscal year. Since that time collections for oil have ranged between $239 million and $354 million, with the exception of fiscal year 1990–91, when collections rose to just over $412 million because of the run-up in oil prices associated with the Persian Gulf conflict.

Natural gas severance tax collections are smaller than oil collections despite the fact that Louisiana produces much more gas than it does oil. Unlike oil severance taxes, which have been tied to oil's price, natural gas severance tax collections have been closely associated with production. In 1972 natural gas was taxed at the rate of $0.033 per thousand cubic feet (mcf). That rate grew to $0.07 per mcf in 1974 and then to $0.10 per mcf in 1990. Since that time natural gas prices have, on net, moderated somewhat, but in July 1997 the rate was set at $0.101 per mcf (Louisiana Department of Revenue and Taxation 1997, 151). The share of total severance tax collections from gas severance tax collections became significant during the early 1960s and grew through the 1970s as marketable uses for natural gas were established and grew. During the late 1970s collections for natural gas approached those of oil, but in the early 1980s surging world oil prices and moderate levels of gas production combined to push natural gas severance taxes' share of total collections down.

Conclusion

Because new technologies are making it easier to identify potential energy deposits and more feasible to extract oil from existing formations and are also enhancing overall efficiency, the average production cost of oil and gas extraction is being driven down. Importantly for the Gulf of Mexico, this beneficial effect of the new technologies becomes relatively more valuable as the geology becomes more complex. That is, extraction costs will fall more dramatically in the gulf, where plentiful oil and gas deposits are found in complex formations, and relatively less in geologically less complex regions like the Middle East, where optimally positioning a well will save some money but the extraction process more resembles draining a large underground pool. In the most extreme cases in the Gulf of Mexico, the net effect of the new technologies is to move oil recovery rates from around 30 percent of potential to over 80 percent.

This analysis suggests some good things for the state of Louisiana, which is the first in line to benefit from these positive industry developments. The current boom is not being fueled by an external rise in product price but rather by a technologically induced reduction in costs. Since the shock to technology will not disappear, the employment and associated income gains seen recently appear likely to last. As a result, the Gulf of Mexico has some immediate gains relative to other, simpler, geologies. Given the state's current taxation policies, however, its tax revenue gains will be limited, coming mostly from rising personal and corporate income because severance revenue is dependent upon the price of oil rather than its profitability (although that may be captured in part by income taxes). This tax structure is similar to other states' and is not necessarily a bad thing, in particular since it looks like gains may be very long term.

The state taxes oil at a higher rate than natural gas, although for mostly geophysical reasons Louisiana produces roughly twice as much gas as oil.

This news comes at a particularly fortuitous time for Louisiana. In the early 1990s, prior to the resurgence in drilling, the state's economy had seen particular strength in tourism and gambling-related construction, and state government had received an increasing share of its revenue from social medical insurance funds. Gambling did not work as well for Louisiana as it did for Mississippi, and most of the growth in, and income from, gambling-related service and construction jobs in Louisiana was short-lived. At the same time tourism, especially in the New Orleans area, grew to its capacity in the mid-1990s, hence limiting further revenue growth. While state tax revenue collection growth was slowing, at the federal level Louisiana was being pressured to reduce the size of its Medicaid expenditures (State Policy Research). These trends, combined with slowing in some previously hot areas of the private sector, made the effects of the technological revolution in the oil industry—particularly its apparent ability to last at current oil prices—a welcome development.

4. The boundary between federal and state waters is typically about three and a half miles from shore.

For the world, the effects of this technology shock have yet to be felt, particularly its implications for offshore energy development. The Gulf of Mexico and Brazil account for almost three-fourths of all deep-water drilling, and the North Sea is rapidly emerging as a significant source. The remainder of the world's deep-water oceans remain unexplored, and the most dramatic reductions in exploration costs are likely to be seen in these untapped regions. Momentum is gathering for deep-water exploration off west Africa, northwestern Europe, and in the Pacific off the coast of several Asian countries ("Deepwater" 1996).

Judging from the case of Louisiana, the other oil- and gas-producing regions will be better off too. The new technology permits greater production at prevailing energy prices. In turn, Louisiana has experienced a very solid base of employment in the relatively high wage energy extraction and related industries. And this surge in employment may not be as fragile as during the OPEC embargoes, when considerable international collusion was required to keep prices up. At the same time, however, this surge in income may have less fiscal impact depending on regions' systems for taxing energy.

REFERENCES

Bennett, Paul, Harold Cole, and Steven Dym. 1980–81. "Oil Price Decontrol and Beyond." Federal Reserve Bank of New York *Quarterly Review* 5 (Winter): 36–41.

"Deepwater Experience, Prospects Outlined." 1996. *Oil & Gas Journal* 94 (September 9): on-line at www.ogjonline.com.

Louisiana Department of Natural Resources. 1996. *Louisiana Energy Facts Annual.*

———. 1997. "Louisiana Energy Topics: Revenue." *Louisiana Energy Facts* (July): insert.

Louisiana Department of Revenue and Taxation. 1997. *1997-1998 Tax Exemption Budget*. Baton Rouge, La.

"Milestones and Influences in U.S. Offshore History (1947–1997)." 1997. *Offshore Magazine* 57 (May 1): on-line at www.ogjonline.com.

Neff, D.B., and T.S. Thrasher. 1993. "Technology Enhances Integrated Teams' Use of Physical Resources." *Oil & Gas Journal* 91 (May 31): on-line at www.ogjonline.com.

"1947—Shaking the Bounds of Land; 1997—Probing 10,000 ft Depths." 1997. *Offshore Magazine* 57 (May 1): on-line at www.ogjonline.com.

Pulsipher, Allan G. 1990. "A Refinery Tax for Louisiana?" Louisiana State University, Center for Energy Studies, May.

Pulsipher, Allan G., Robert H. Baumann, and Wumi O. Iledare. 1993. "Overall Tax Levels Similar across 11 Producing States." *Oil & Gas Journal* 91 (October 11): on-line at www.ogjonline.com.

State Policy research, Inc. 1995. *State Budget & Tax News* 14 (November 28): 5–6.

A Primer on Short-Term Linkages between Key Economic Data Series

R. MARK ROGERS
The author is the forecast coordinator in the macropolicy section of the Atlanta Fed's research department.

IN THE UNITED STATES, ECONOMY WATCHERS ARE BLESSED (OR CURSED, DEPENDING ON ONE'S VIEWPOINT) WITH A PLETHORA OF DATA. TO THE CASUAL OR NEW OBSERVER OF THE ECONOMY, THE INFORMATION CONTENT OF THE MANY INDICATORS MAY BE UNCLEAR. MOREOVER, THE WAY EXPERIENCED ANALYSTS USE THE DATA TO EVALUATE THE ECONOMY MAY SEEM COUNTERINTUITIVE. TO THE NEWCOMER THE QUESTION OFTEN IS, WHY DO THOSE INTERESTED IN FUTURE ECONOMIC DEVELOPMENTS LOOK AT THESE NUMBERS? OR HOW DOES ONE USE THESE DATA SERIES? THERE MAY BE AS MANY ANSWERS AS THERE ARE ANALYSTS.

Why do analysts look at economic data? The simple answer is that investors and planners must look forward, and economic data help them forecast. If there is new information on the economy, on demand, on profit potential, or on prices, among other factors, then the underlying value of financial and real investments may shift, changing values, project projections, and plans.

The release of economic data can have an impact on the value of financial instruments and investment projects because it may change analysts' views of the strength or weakness of the economy. These views in turn may affect their forecasts for company or project earnings, general or specific prices, and interest rates. Because major decisions may depend on economic reports, market participants need to squeeze as much information as possible out of data so as to make intelligent decisions about financial holdings and investments.

There are a number of time horizons relevant to how market watchers evaluate economic data and use them for forecasting. The evaluation of various longer-run fundamentals often begins with examining short-run relationships among economic variables. This article focuses on these relationships—many of which involve lagged effects taking place over a few months or at least within a year and a half. Clarifying the source data linkages and the statistical linkages will help explain how and why financial markets track and react to economic data the way they do. Source data are series from one statistical agency used by another statistical agency to derive a new series (discussed below).

This article is a brief guide to some of the well-known short-term relationships between economic data series upon which many analysts focus. It explains how analysts use data in concurrent month forecasts and what some key relationships are, outlines the monthly calendar of economic releases, and, finally, reports on typical lags between various dependent and explanatory variables.

Concurrent Month Linkages—Source Data and Statistical Links

There are at least two basic approaches to linking two or more economic series over a short time horizon. Analysts try to use prior-released data to project later-released data for the same period. This practice is called "forecasting" concurrent data. Linkages between the earlier and later data sets may be based on common source data or on some statistical relationship.

Source Data. Source data are series used by a statistical agency (usually a government bureau such as the Commerce Department's Bureau of Economic Analysis [BEA]) to derive other economic statistical series. For example, the BEA uses average hourly earnings data from the Labor Department to help produce the wages-and-salaries component in personal income data; the BEA also uses residential construction outlays to help estimate the residential investment component of gross domestic product (GDP).

Analysts should be aware that analyzing source data to forecast concurrent data for the derived data series is less than straightforward because the statistical agencies typically make numerous adjustments to the source data at levels of detail not always accessible to the public. Adjustments may have been made for differences in definition, geographic coverage, or timing and obvious or subtle differences in economic concepts. Additionally, a subsequent data series may be based on more than one set of source data. Some examples follow.

The industrial production index has three principle components: manufacturing production, utilities production, and mining. For the initial release of the index, the Federal Reserve Board of Governors bases the manufacturing production component primarily on production worker hours in the manufacturing sector, available from the establishment survey in the employment report produced by the Bureau of Labor Statistics (BLS). The Federal Reserve Board uses this measure as its primary input for the initial estimate because so little hard data for actual production are available for the month about to be released. (For example, data on kilowatt hours of electricity used in production are not available until later in the month. For later revisions to the initial estimates of industrial production, the Federal Reserve Board incorporates these other types of data.) The production worker hours data become publicly available on the first Friday of each month following the reference month; the industrial production report for the same reference month is released around the fifteenth of the following month. Thus, on average, the production worker hours data are available about ten days prior to the production index release. Analysts use this data to judge the strength of the manufacturing sector in terms of estimated output. As a percentage of value added in 1994, production worker hours data underlay 29.1 percent of the initial estimate directly and 53.1 percent indirectly (for heavily judgmentally based series). These percentages, respectively, were 29.1 percent and 2.5 percent for the fourth month estimates.[1]

The evaluation of various longer-run fundamentals often begins with examining short-run relationships among economic variables.

Because production workers hours data are key inputs for initial estimates of industrial production, market analysts often attempt to forecast an upcoming release with a regression model based on production worker hours data, shown in Table 1.[2] This regression estimates manufacturing output as a function of production worker hours plus a constant over the 1980–97 period. Both output and hours are in monthly percentage change form. The percentage change in manufacturing output is estimated to be 0.631 times the percentage change in production worker hours plus a constant of 0.268. This simple model has reasonably good explanatory power with an adjusted R^2 of 0.5015 and with t-statistics for both explanatory variables statistically significant. (R^2 is the coefficient of determination, a statistical measure of the "explained" variation in the data as a percentage of the total variation in the data. Values for R^2 range from 0 to 1.00 so that, for a simple regression model with only one explanatory variable, all the data lie on the regression line when R^2 equals 1.00—that is, there are no unexplained

1. *These figures are unpublished estimates by Federal Reserve staff, October 1997.*
2. *This particular regression model is discussed in more detail in Rogers (1992). Importantly, one should note that the Federal Reserve Board estimates production with procedures for individual components. About 82 percent of the series is based on production worker hours, directly and indirectly, for the initial estimate. Even for these series, the Federal Reserve Board makes estimates using production factor coefficients (PFCs) based on more than just production hour data (see Board of Governors 1986, 33–128).*

   PFCs are used to estimate individual industrial production series, which are estimated over historical periods, taking into account trend and cyclical relationships between production and the hours input and adjusting the hours data to be representative of the month as a whole. This procedure is more complex than is represented by a simple production hours regression model, but market analysts have found this type of model to have some usefulness.

TABLE 1 A Simple Model for Forecasting IP with Production Worker Hours

Regression using OLS
Dependent variable: FRB industrial output, manufacturing, percent change
Explanatory variable: Production worker hours, percent change
1980M1–97M10

Variable	Coefficient	Standard Error	*t*-statistic	Significance
Constant	0.267801	0.380935E-01	7.03011	0.000
Production Worker Hours, Percent Change	0.631322	0.430306E-01	14.6715	0.000

Equation Summary

Number of Observations	=	214	R^2 (adjusted)	=	0.5015
Sum of Squared Residuals	=	65.6672	Standard Error of Regression	=	0.556553
R^2	=	0.5038	Durbin-Watson	=	2.36697

variations in the data. Adjusted R^2 is a measure that takes into account how many explanatory variables are used in the regression model.) Based on its moderately high adjusted R^2, the regression confirms that a percentage change in manufacturing production worker hours is useful for forecasting manufacturing output for the current month.

Statistical Relationships. A second way that analysts may make forecasts—short-term or long-term—is by linking different data series that have relatively dependable statistical relationships. Two data series may have a common near-term link either to each other or to separate variables even though statistical agencies do not use one series to produce the other. An example of such a concurrent-month—or same-reference-month—statistical relationship is using producer price index (PPI) data, released earlier in the month, to project the consumer price index (CPI), even though the BLS derives these indexes independently.

Another example involving short-term linkages of data for concurrent forecasting is using the purchasing managers index to predict the later-released industrial production index.[3] Even though the former index is not used to produce the latter, there is a statistical relationship between the direction and magnitude of movement in the purchasing managers index and the industrial production index on a concurrent month basis. The National Association for Purchasing Management (NAPM) releases a survey of manufacturers in its association on the first business day of each month following the reference month. This release is timed so that manufacturing sector data are available on average about two or three days before the employment situation data on production worker hours. This early release date for manufacturing sector data makes this release a very important one for profit-driven analysts, who are motivated to determine whether it contains any significant information that will help them assess the strength of the economy before the employment report is released.

The purchasing managers survey release contains a composite index, the components of the composite index, and a number of indexes not included in the composite. The composite index is based on subcomponents for production, new orders, employment, inventories, and vendor performance. To predict the release of industrial production later in the month—but prior to the release of the employment situation—analysts typically regress the percentage change in industrial production against the NAPM composite diffusion index. This diffusion index measures not levels of activity but percentages of respondents indicating an increase, decrease, or no change in activity. NAPM's diffusion index is the percentage of respondents indicating an increase in activity plus half the percentage indicating no change. Hence, the level of these diffusion indexes is associated with percentage changes in corresponding government data series based on actual dollar values or output level. For this article's statistical comparison, the Federal Reserve Board's manufacturing output index in monthly percent changes is regressed against the NAPM's production diffusion index level. This regression model, shown in Table 2, estimates that the percentage change in manufacturing output is equal to 0.055 times the NAPM production index plus a constant of –2.758. The explanatory power of this model, with an adjusted R^2 of 0.2991, is lower than the production worker hours model (Table 1), but analysts use this type of model because the NAPM data are released prior to the production worker hours data and the NAPM data's explanatory power is significant.

Statistical relationships can be expanded beyond the current month when one variable "explains" a second

TABLE 2 A Simple Model for Forecasting IP with the NAPM Production Index

Regression using OLS
Dependent variable: FRB industrial output, manufacturing, percent change
Explanatory variable: NAPM production diffusion index level
1980M1–97M10

Variable	Coefficient	Standard Error	*t*-statistic	Significance
Constant	–2.75805	0.315984	–8.72845	0.000
NAPM Production Index	0.554479E-01	0.578480E-02	9.58510	0.000

Equation Summary

Number of Observations	=	214	R^2 (adjusted)	=	0.2991
Sum of Squared Residuals	=	92.3292	Standard Error of Regression	=	0.659936
R^2	=	0.3023	Durbin-Watson	=	2.21014

series over an extended time horizon (even if only for a few months). One series in a base time period typically has some known economic impact on another series in a subsequent time period. For example, changes in housing permits over time lead to changes in housing construction outlays.

In summary, short-term analysis of data can involve concurrent forecasting using either source data or well-known statistical relationships among explanatory variables. The use of independent variables can be expanded beyond current period analysis to longer-term forecasting.

Monthly Releases and Concurrent Linkages

Analysts' abilities to predict economic strengths as much in advance as possible depend on the fact that there is a regular cycle to economic news releases. Federal government statistical agencies typically give dates for economic news releases for a given year during the latter part of the previous year. The relative order of each release during the calendar month has changed little over the years. For example, the U.S. Department of Labor generally releases the employment situation report on the first Friday of each month. Industrial production is usually released by the Federal Reserve Board of Governors around midmonth, and GDP estimates typically are released during the last week of each month. Other government—and private-sector—release dates are also generally known well in advance and have followed much the same sequence relative to each other for years. For example, the PPI always precedes the CPI, usually by about three days. Table 3 gives a typical schedule of key economic releases over a monthly release cycle.

What series are used to project subsequently released concurrent month data? And what are the basic relationships between the released and the projected series? Table 4 lists the primary linkages for concurrent month forecasting according to when key data series are first made public. Series in the left-hand column are released to the public prior to those in the right-hand column. Table 5 shows the primary source data specifically for GDP components. Financial markets track economic series in the sequence that they are released publicly. The key reports shown in Tables 4 and 5 and linkages from those reports to later-released data are discussed below.

The Purchasing Managers Report. Several individual series from the monthly report by the National Association of Purchasing Managers are used to predict other, later-released economic data. Although the most notable instance is the use of the purchasing managers production index to predict the industrial production index produced by the Federal Reserve, discussed earlier, there are others. Some analysts use the NAPM composite index rather than the production index as the explanatory variable. Other series are used to a lesser degree because the statistical relationship is less reliable. The NAPM employment index is used to predict BLS data for nonfarm payroll employment—or, more specifically, for the manufacturing employment component of the establishment employment report. The NAPM prices paid index is often correlated with the BLS producer price index. The NAPM new orders index has a small predictive capability for the Census Bureau's new factory orders. Finally, the Conference Board uses the NAPM vendor performance index as source data directly

3. For more detailed discussion of these types of models, see Rogers (1988, 1992, 1994, and 1998), Harris (1991), and Harris and Vega (1996).

TABLE 3 Monthly Release Schedule for October 1997

Release Date	Indicator	Reference Period
October		
1	Construction expenditures	August
1	Purchasing managers index, NAPM	September
1	Conference Board's composite indicators	August
2	Manufacturers shipments, inventories, and orders	August
2	Initial unemployment claims	September 25
3	Employment situation	September
6	Auto sales, AAMA	September
8	Wholesale trade	August
9	Initial unemployment claims	October 4
10	Producer price index	September
14	Atlanta Fed manufacturing survey	September
14	Richmond Fed manufacturing survey	September
15	Advance monthly retail sales	September
16	Consumer price index	September
16	Initial unemployment claims	October 11
16	Philadelphia Fed manufacturing survey	October
16	Business inventories and sales	August
17	Housing starts and permits	September
17	Industrial production and capacity utilization rate	September
21	U.S. international trade in goods and services	August
23	Initial unemployment claims	October 18
28	Employment cost index	Third Quarter
29	Advance report on durable goods	September
30	New one-family house sales	September
30	Initial unemployment claims	October 25
31	GDP	Third Quarter
November		
3	Personal income, outlays, and saving	September
3	Purchasing managers index, NAPM	October

for that component in the Conference Board's index of leading indicators.

The Employment Situation Report. The employment situation report, released the first Friday of each month after the reference month, contains four major sets of data series used for concurrent month forecasting. The report's primary importance stems from the fact that it is the first major release each month with comprehensive coverage of all major sectors of the economy; the report provides key data on the strength of the manufacturing and consumer sectors. As already discussed, the manufacturing production worker hours index is used by the Federal Reserve Board to estimate the first release figure for manufacturing output. Second, the BEA uses nonfarm payroll data on employees, the average workweek, and average hourly earnings to estimate the private-sector portion of wage and salary disbursements in the personal income report. Next, the manufacturing average workweek is one of the components of the Conference Board's composite index of leading indicators. Finally, the series for nonfarm payroll employment is part of the Conference Board composite index of current indicators.

American Automobile Manufacturers Association (AAMA). The AAMA, formerly known as the Motor Vehicle Manufacturers Association, produces data on unit sales for autos and light trucks. The BEA uses these data to estimate portions of GDP components—notably durables personal consumption expenditures, producers durable equipment, and government consumption expenditures and gross investment. These components reflect purchases or leases of light motor vehicles.

Chain Store Sales—LJR Redbook. Several private firms produce reports on weekly or monthly chain store sales. The most widely known is the weekly series produced by the New York investment firm of Lynch, Johnson, and Ryan, published in their Redbook report. (This report was previously called the Johnson

TABLE 4 Indicators for "Forecasting" within the Monthly Cycle

Precursor/Explanatory Series and Producing Agency	Series Being "Forecast" and Producing Agency
Purchasing Managers' Report, NAPM	
(a) composite or production index	(a) industrial production, FRB
(b) employment index	(b) manufacturing employment, BLS
(c) prices paid index	(c) producer price index, BLS
(d) inventory index	(d) manufacturers inventories, Census
Employment Report, BLS	
(a) aggregate production hours in manufacturing[a]	(a) industrial production, FRB
(b) average hourly earnings, payroll employment, average workweek[a]	(b) wage and salary disbursements in personal income report, BEA
(c) average manufacturing workweek[a]	(c) component of index of leading indicators, Conference Board
(d) nonfarm payroll employment[a]	(d) component of index of current indicators, Conference Board
Unit New Auto Sales, AAMA	
(a) auto and light truck sales[a]	(a) durables PCEs in personal income report, BEA
LJR Redbook	
(a) chain store sales	(a) department store sales in retail sales report, Census
Retail Sales, Census	
(a) retail sales[a]	(a) durables and nondurables PCEs in personal income report, BEA
Producer Price Indexes, BLS	
(a) consumer product components	(a) goods components in CPI, BLS
Manufacturers' Shipments, Inventories, and Orders, Census	
(a) nondefense capital goods shipments[a]	(a) producers' durable equipment in GDP, BEA
(b) manufacturers inventories[a]	(b) change in inventories, manufacturers, in GDP, BEA
Monthly Business Inventories, Census	
(a) business inventories[a]	(a) inventory change in GDP, BEA
Monthly International Trade, Census and BEA	
(a) goods and services exports and imports[a]	(a) net exports in GDP, BEA
Construction Outlays, Census	
(a) residential outlays[a]	(a) residential investment in GDP, BEA
(b) nonresidential outlays[a]	(b) nonresidential structures in GDP, BEA
(c) public outlays[a]	(c) structures component in government purchases in GDP, BEA

[a] Source data for forecast series

Note: FRB indicates Federal Reserve Board of Governors; BLS indicates Bureau of Labor Statistics; BEA indicates Bureau of Economic Analysis.

TABLE 5 Principal Source Data for GDP: Availability for the Advance GDP Release

GDP Component and Monthly Series	Months Available
Personal Consumption Expenditures	
Retail sales	3
Unit auto and truck sales	3
Nonresidential Fixed Investment	
Unit auto and truck sales	3
Value of construction put in place	2
Manufacturers' shipments of machinery and equipment	2
Exports and imports of machinery and equipment	2
Residential Investment	
Value of construction put in place	2
Housing starts	3
Change in Business Inventories	
Manufacturing and trade inventories	2
Unit auto inventories	3
Net Exports of Goods and Services	
Merchandise exports and imports	2
Government Consumption Expenditures and Gross Investment	
Federal outlays	2
Value of construction put in place by state and local government	2
GDP Prices	
CPI	3
PPI	3
Nonpetroleum merchandise export and import price indexes	3
Values and quantities of petroleum imports	2

Redbook, named after the individual who started the report.) The LJR chain-store data are compiled from public reports from major chain stores in the United States. The weekly data, which are released on Tuesday afternoons, are not source data for any government statistics on retail sales. But analysts take an interest in the LJR Redbook data because they are available prior to the Commerce Department's retail sales report, are somewhat indicative of the strength of consumer spending, and have moderate predictive power for the narrowly defined department store series within the retail sales report.

The Retail Sales Report. Commerce's report on retail sales is released around midmonth following the reference month. The Census data on retail sales are used by the BEA to produce estimates for portions of personal consumption expenditures, which are part of GDP. The retail sales data are also used in the "disposition of income" portion of the Personal Income report, which is released the next business day after GDP estimates. Markets look at the retail sales data because they are a major indicator of consumer strength and they precede the personal consumption numbers by about two weeks. However, retail sales do not cover services and as such are only source data for durables and nondurables portions of personal consumption expenditures (PCEs). (In 1997, durables and nondurables PCEs were 12.0 percent and 29.0 percent, respectively, of total nominal PCEs. The BEA uses AAMA data for motor vehicle consumption because those numbers are more reliable than the Census survey–based data for retail sales. The AAMA data essentially cover all sales as tallied by the auto manufacturers themselves.)

Producer Price Index. The producer price index is released midmonth following the reference month. It precedes the CPI report by about three days, and analysts use the PPI numbers to project the CPI release figures. PPI data are not source data for CPI data; the data sets are derived from two independent surveys. The predictive power of PPI data for CPI numbers is only moderately strong, as suggested by standard regression statistics (see Rogers 1988). There are some notable definitional differences between the PPI for finished goods and the all-urban CPI. For example, the PPI does not cover services but does cover capital equipment; about half of the CPI component weight is services, but the CPI does not cover capital equipment. Also, even for components that are very similar for the PPI and CPI, such as food and energy, the rate at which prices at the producer level pass through to the consumer level varies by component.

Manufacturers Inventories, Orders, and Sales. This report, produced by the Census Bureau, contains source data for two components of GDP. The manufacturers inventories data from the monthly Census report form the backbone of the manufacturers component of inventory investment within GDP. However, the relationship is not as tight as might be expected because the BEA must make substantial adjustments in the Census data to convert them to the proper form for National Income and Product Accounts (NIPA).[4]

The monthly manufacturers report also provides source data for a second GDP component: producers durable equipment. Analysts focus on data for nondefense capital goods equipment shipments within the orders report as a barometer of future spending on producers durable equipment. But the relationship of nominal shipments of nondefense capital goods shipments with nominal producers durable equipment investment is not as tight as might be expected. The relationship is not one-for-one for two primary reasons: not all capital equipment produced in the United States is sold to domestic users, and U.S. businesses obtain capital equipment not only from domestic producers but also from those overseas. Therefore, in the manufacturers inventories, orders, and sales report, exports of capital equipment are subtracted from domestic equipment investment—that is, producers durable equipment and imports of capital equipment are added, but the latter are not part of (domestic) shipments of nondefense capital goods as measured in the Census report.

Monthly Business Inventories. The business inventories report is a later-published, broad report on overall business inventories. It includes the earlier-released manufacturers inventories plus data for merchant wholesale inventories and retail inventories. These data are source data for nonfarm inventory investment within the GDP accounts. As with the manufacturers data, there are a number of adjustments made by the BEA in converting the wholesale and retail series to their NIPA equivalents.

Monthly International Trade. Monthly international trade data, jointly produced by the Census Bureau and the BEA, are source data for goods and services exports and imports in the GDP accounts as well as in the balance of payments accounts. There are a notable number of coverage and timing differences between the monthly series and the balance of payments series and, in turn, the GDP series. One coverage difference is that the customs data that go into Census data are based on the geographic authority of U.S. Customs, which includes U.S. territories. Data that include U.S. territories are appropriate for balance of payments data but are not appropriate for GDP accounts within NIPA since GDP is defined by national borders exclusive of territories.

Short-term analysis of data can involve concurrent forecasting using either source data or well-known statistical relationships among explanatory variables.

Monthly Construction Outlays. Monthly construction outlays data, or construction spending data, produced by the Census Bureau, are key source data for various structures components within GDP. Monthly construction spending data serve as a measure of production in the construction sector. Data on private residential outlays are source data for GDP's residential investment component; nonresidential outlays, for nonresidential investment; and public construction outlays, for structures components within government consumption expenditures and gross investment. The statistical relationship between these series is moderately strong, based on regression analysis, because the monthly outlay series source data are not the only source data used for GDP structures components. Additional source data includes for example, a quarterly survey used to estimate spending on additions and

4. The National Income and Product Accounts, produced by the BEA, are broad "double-entry" accounts that track economic activity in the United States. With double-entry accounts, for every expenditure series there is a corresponding income account; the NIPA accounts attempt to follow economists' definition that spending generates an equal amount of income. For GDP estimates based on expenditures (such as personal consumption and investment, among others), there are GDP estimates based on personal income, corporate profits, and other income components.

alterations, which are part of the GDP residential investment series, and a subcomponent for brokerage commissions.

Key Source Data for GDP. Analysts project GDP ahead of its official release because it is viewed as a summary measure of overall economic performance. Tracking various releases for source data is important for developing an estimate for current-quarter GDP as the release months of the quarter progress. Most of the key series of GDP source data are listed in Table 5, which pulls together many of the series listed by separate reports. A more complete listing is available from the U.S. Commerce Department (1996).

One key difference between estimating GDP from earlier-released source data and using source data to estimate other monthly series is that one or more months of data are missing for some component series when the first release for GDP is made to the public. Quarterly GDP is revised each month for two months after the initial release. The first release is referred to as "advance;" the second, as "preliminary;" and third, as "revised." Table 5 shows how many months of data are available for each source data series when GDP is initially released for a given quarter. The BEA also uses additional unpublished data that may not be available to the public. In addition, some source data may not be available at all for the early estimates of current-quarter GDP and become available only by the time of the annual revisions during the subsequent year. In these cases the services components are projected for the current quarter since they are derived from private-sector annual surveys.

One series in a base time period typically has some known economic impact on another series in a subsequent time period.

When the BEA releases the advance estimate for GDP, it also publishes its assumptions for missing months of data for monthly source data that are public. This table, titled "Summary of Major Data Assumptions for Advance Estimates," is published in the *Survey of Current Business* with the advance GDP report. Comparing subsequent releases of missing monthly data with the BEA's assumptions provides some clue toward the direction of later revisions to current-quarter GDP. However, because monthly source data are only one part of the estimation procedure, differences between BEA assumptions and subsequent releases provide only part of the explanation for subsequent revisions to GDP estimates.

Behavioral Links between Data Series

Analysts use economic data to forecast other economic series by observing various behavioral links. That is, one type of economic activity appears to have an impact on another type of economic activity, and often with a lag. For example, a rise in factory orders is believed to lead to an increase in industrial production. Although a detailed explanation of econometric models for various sectors in the economy is beyond the scope of this article, a brief discussion of some basic behavioral linkages between economic data series and what type of lagged impact one variable has on the other will round out this primer on data series.

The Consumer—Income and Expenditures. An income-expenditure flow analysis of the consumer sector is relatively straightforward. Income is the "driver" behind consumer spending, although other factors play a role. Additional fundamentals include changes in employment and wealth, changes in interest rates, and changes in prices. Nonetheless, a key to understanding the consumer income-expenditure flow is to examine what determines—in simple terms—consumer income. Aggregate consumer income, in a definitional sense, is based on the product of the number of workers, the average number of hours worked, and the average wage. The data series that correspond to these concepts are nonfarm payroll employment, the nonfarm average workweek, and average hourly earnings. All of these series are part of the employment situation report produced by the BLS and form the backbone of the BEA's estimates of the wage and salary disbursement portion of personal income. Analysts track these series in part so that they can gauge the strength of consumers' ability to spend.

As the flow diagram in Chart 1 shows, an increase in either employment, the average workweek, or average hourly earnings leads to an increase in personal income, and, in turn, an increase in personal consumption. Of course, this flow assumes that all other factors are held constant as the factors in the behavioral flow change. Clearly, other factors come into play in determining consumer spending, but in this simplified model these outside factors have no impact on explaining changes in consumer spending. Similarly, as the article discusses other behavioral flows between economic data series, for variables not discussed, the assumption of ceteris paribus is made.

Manufacturing and the Inventory Cycle. Income and expenditures flows play a more complex role in the manufacturing sector in what is traditionally called an inventory cycle. Essentially, changes in consumer spending affect actual and desired inventory levels; when desired inventory levels differ from actual levels, manufacturers, wholesalers, and retailers make necessary adjustments to bring the two together. These actions, in

CHART 1 Behavioral Flow for Consumer Spending

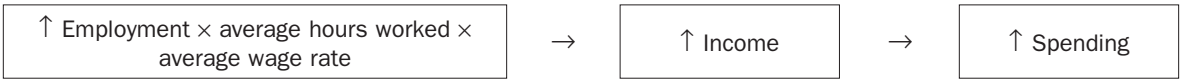
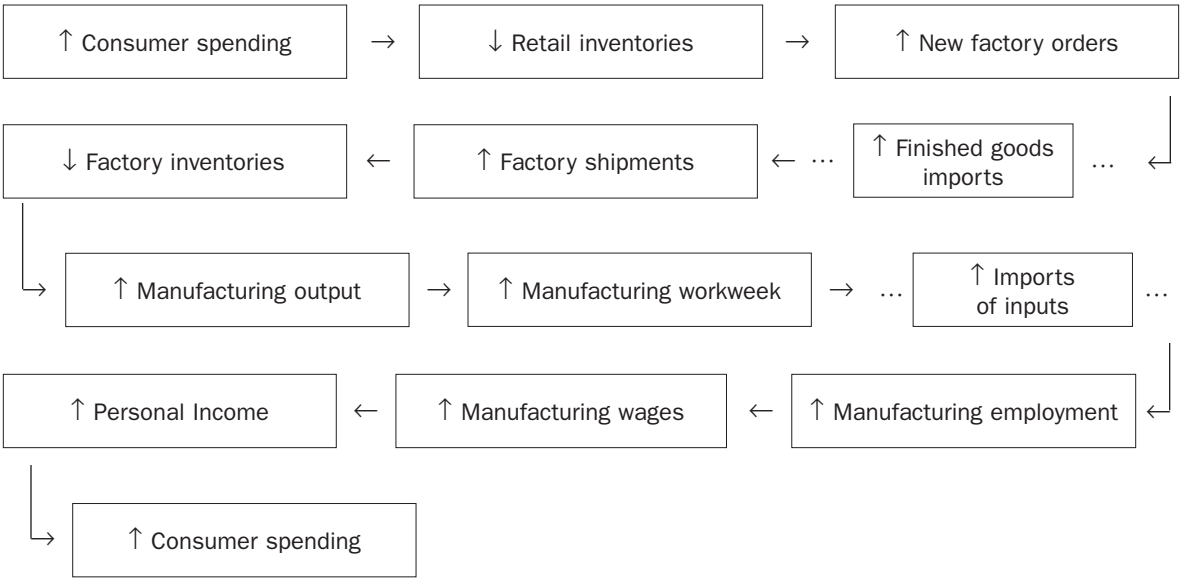


CHART 2 Behavioral Flow for Manufacturing



turn, affect consumer income and spending. Essentially, the consumer plays a key role in the inventory cycle.

The flow diagram in Chart 2 illustrates this cycle. At the beginning of the cycle, if retail sales to consumers are unexpectedly strong, retail inventories will decline below desired levels. Retailers will then place new orders with domestic producers or order additional imports. Domestic producers respond by increasing shipments, which in turn cause manufacturers' inventories to drop below desired levels. This drop in turn boosts manufacturing output. Initially, manufacturers merely increase the average workweek, but when demand is sufficiently strong they hire additional workers. Average hourly earnings may rise in order to attract the additional workers. As personal income rises, the cycle is renewed because this income gain can fuel additional consumer spending.

Table 6 lists the data series that analysts track to follow this cycle. The left-hand column shows the generalized economic concepts in the behavioral flow for manufacturing while the right-hand column indicates the specific data series that correspond to the economic concept.

Analysts are interested in determining the average length of time it takes for a change in one variable to affect a second variable. But a complicating factor in estimating some of these average lag lengths is that the direction of causality is not always consistent, especially for sales series and inventory data. During a business cycle, businesses may do a better job of anticipating sales at some times than at others; inventory changes may anticipate sales changes and vice versa. This reciprocity reduces the likelihood that measures of average lag length are statistically meaningful for these data series. For other data series, the causal relationships may be more consistent so that average lag lengths can be estimated. For example, housing starts essentially always precede housing outlays, and changes in durables factory orders precede changes in durables production.

For manufacturing sector analysis, it is useful, for a couple of reasons, to segment the discussion between durable goods and nondurable goods. First, durable goods, especially heavy capital equipment, tend to have longer production cycles than nondurables, and durables output is more cyclically sensitive. Changes in durables and nondurables share of output over the business cycle would affect the reliability of estimates of various lag coefficients (such as orders to production) if estimated using data that were not disaggregated between durables and nondurables. Second, differences in methodologies for nondurables orders affect lag estimates.

TABLE 6
Behavioral Flow for Manufacturing: Economic Concepts and Corresponding Data Series

Economic Concept	Data Series
Consumer spending	Retail sales, Census Personal consumption expenditures, BEA Unit new motor vehicle sales, AAMA and BEA Chain store sales, LJR
Retail inventories	Retail inventories, Census
New factory orders	Manufacturers new orders, Census
Factory shipments	Manufacturers shipments, Census
Factory inventories	Manufacturers inventories, Census
Imports	Imports of goods and services, BEA and Census
Manufacturing output	Industrial production index, Federal Reserve Board Manufacturing surveys: NAPM, Chicago Purchasing Managers, Atlanta Fed, Kansas City Fed, Philadelphia Fed, and Richmond Fed
Manufacturing workweek	Average workweek, manufacturing, BLS
Manufacturing employment	Nonfarm payroll employment, manufacturing, BLS
Manufacturing wage	Average hourly earnings, manufacturing, BLS
Personal income	Personal income, BEA

Table 7 shows the average (mean) lags, estimated by the Almon distributed lag technique, between various manufacturing sector indicators for durables. The mean lag between changes in new factory orders for durables (real) and industrial production for durables manufactured goods in all durables industries is 2.3 months;[5] the lag from production to shipment is relatively short—only 1.654 months. Lags would vary if they were estimated on an industry-by-industry basis. For example, the orders-to-production mean lag would be much longer for the aircraft industry than for the lumber industry. Estimates of lag length also vary depending on the model and lag structure chosen (see Greene 1993, 519–25).

For nondurables, data methodology for orders has an interesting impact on lag estimates. The Census Bureau's monthly estimates for new orders are defined as current-month shipments plus current-month unfilled orders minus prior-month unfilled orders (see Rogers 1994, 145). This formula works reasonably well for industries with unfilled orders. However, most nondurables industries report no unfilled orders—for December 1997 only 25.5 percent of the dollar value of new orders for nondurables was for industries that report unfilled orders. For industries with no unfilled orders, Census uses shipments data for new orders—that is, new orders are assumed to equal the available shipments numbers and to represent post-production activity. Official data indicate that most nondurables production takes place during the same month as the shipments/new orders. Table 8, which reports on the regression output of nominal shipments regressed against contemporaneous nominal new orders and a constant, shows the high correlation (an adjusted R^2 of 0.9378) between nondurables new orders and nondurables shipments.

Construction Sector Linkages. Just as there are inventory cycle effects in manufacturing, there are similar linkages in the construction sector (see Table 9). An unexpected increase in housing sales leads to a drop in houses for sale as well as in the months' supply of houses for sale. Houses for sale and months' supply are the housing sector's equivalent of manufacturers' inventories data and of the inventories-to-sales ratio. If hous-

TABLE 7 Manufacturing Indicators: Lags between Key Series

Estimation technique: Almon distributed lag
Observation period: 1970M1–97M9

Series and Predecessor Series	Mean Lag (Months)	Standard Error of Mean Lag	Adjusted R^2 of Equation	Lag Specification (Order, Lag Length, Endpoint Constraint)
Industrial production, durables/ Durables orders, real	2.342	Undefined	0.401	2, 12, None
Durables shipments, real/ Industrial production, durables	1.654	Undefined	0.309	3, 9, None

TABLE 8
High Correlation between Same-Month Nondurables New Orders and Shipments

Regression: Dependent variable is
nondurables, shipments, nominal, percent change
1970M1–97M9

Variable	Coefficient	Standard Error	*t*-statistic	Significance
Constant	0.320391E-01	0.192568E-01	1.66378	0.097
New Orders, Nominal, Percent Change	0.940697	0.132951E-01	70.7552	0.000

Equation Summary

Number of Observations	=	333	R^2 (adjusted)	=	0.9378
Sum of Squared Residuals	=	34.8928	Standard Error of Regression	=	0.324679
R^2	=	0.9380	Durbin-Watson	=	2.92446

ing stocks decline below desired levels, then builders take out housing permits, initiate housing starts, and work toward completing houses by making construction outlays (spending), as Chart 3 demonstrates. As in manufacturing, this cycle can differ when production is based on expectations of changes in the business cycle. For example, housing stocks may be built up in anticipation of housing sales rather than housing being replenished after a rise in sales. There clearly are times that the direction of causality among some of the inventory-sales-permits-starts linkages reverses, reducing the statistical reliability of these relationships.

Table 10 shows that the average lag (using the Almon distributed lag estimation technique) between changes in housing permits and housing starts is very short—only 1.026 months. The average lag from changes in starts to changes in construction outlays is 4.032 months.

Price Sector Linkages. To some degree there are linkages in prices in various sectors of the economy

5. Because output for durables and nondurables is in real (inflation-adjusted) terms, it is appropriate that the orders and shipments data be converted from current dollars to real dollars. Durables and nondurables orders and shipments data were deflated using BLS data for producer price indexes for durables manufactured goods and nondurables manufactured goods, respectively.

Durables industries include lumber and products, furniture, and fixtures; clay, glass, and stone products; primary metals, fabricated metal products, industrial and commercial machinery, and computer equipment, electrical machinery, transportation equipment, instruments, and miscellaneous manufactures. Nondurables industries include foods, tobacco products, apparel products, paper and paper products, printing and publishing, chemical and products, petroleum products, rubber and miscellaneous plastics products, and leather and leather products.

TABLE 9
Behavioral Flow for Construction: Economic Concepts and Corresponding Data Series

Economic Concept	Data Series
Housing sales	New single-family housing sales, Census Existing single-family housing sales, National Association of Realtors (NAR)
Houses for sale, ratio stocks/sales	Months supply, new single-family houses, Census Months supply, existing single-family houses, NAR
Housing permits	Housing permits, Census
Housing starts	Housing starts, Census
Residential construction spending	Residential construction outlays, Census

TABLE 10 Construction Indicators: Lags between Key Series

Estimation technique: Almon distributed lag
Observation period: 1970M1–97M9

Series and Predecessor Series	Mean Lag (Months)	Standard Error of Mean Lag	Adjusted R^2 of Equation	Lag Specification (Order, Lag Length, Endpoint Constraint)
Housing starts/ Housing permits	1.026	Undefined (Lag signs switch)	0.389	3, 6, None
Residential construction outlays, 1992$/Housing starts	4.032	0.450	0.537	2, 15, None

TABLE 11 Inflation Indicators: Lags between Key Series

Estimation technique: Almon distributed lag
Observation period: 1970M1–97M9

Series and Predecessor Series	Mean Lag (Months)	Standard Error of Mean Lag	Adjusted R^2 of Equation	Lag Specification (Order, Lag Length, Endpoint Constraint)
CPI, total/ PPI, finished goods	1.573	0.144	0.621	3, 6, None
PPI, finished goods/ PPI, intermediate products	0.100	0.105	0.561	2, 4, None
PPI, intermediate products/ PPI, crude materials	5.056	0.460	0.347	4, 12, None

CHART 3 Behavioral Flow for Construction

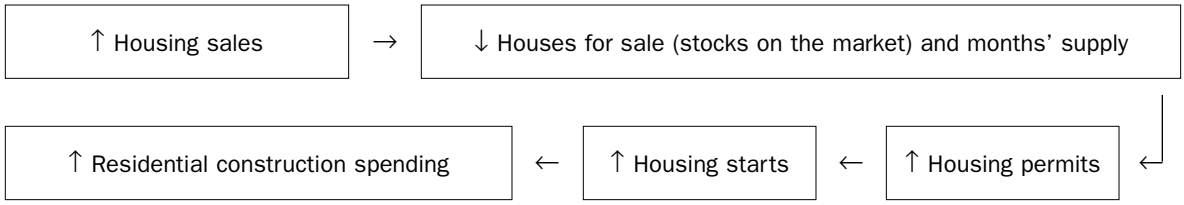
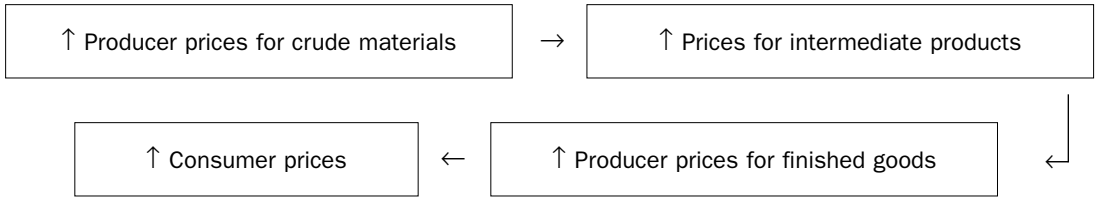


CHART 4 Behavioral Flow for Prices



Note: The individual units in the chart are the names of indexes produced by the BLS. However, producer prices for crude materials refers not only to the index published by the BLS but also to series by the Commodity Research Bureau (CRB) and the *Journal of Commerce* and the prices paid diffusion index from the NAPM.

through cost pass-through. The cost for crude materials may be passed through to costs for intermediate goods, for producer prices for finished goods, and on to the consumer (see Chart 4).

The relationship between the PPI for finished goods and the CPI should be measured using the CPI for goods only (that is, excluding services, since the PPI for finished goods has no services other than electricity from public utilities). In addition, the length of pass-through from the PPI for finished goods to the CPI is rather short, with most of the impact taking place within the current and following months. Finally, the relationship between any two price series above is not particularly strong because there is a great deal of volatility in the data, more so for producer prices for crude materials than for finished goods. Crude materials prices and diffusion indexes provide many false signals of building price pressures at the consumer level. However, rising crude and intermediate prices are generally precursors of an increase in consumer price inflation.

Table 11 shows a very short lag time from changes in producer prices for finished goods and consumer prices—only 1.573 months. Movement in prices for finished goods and intermediate goods is essentially coincident, with an estimated lag of 0.100 month. The apparent pass-through of changes in crude materials prices to intermediate products is somewhat longer, with an estimated mean lag of 5.056 months.

Summary

This article is a primer on some of the key short-term economic relationships among data series upon which economic analysts focus. Certainly, market participants closely watch the calendar of economic releases and, as each release is made, enter the new information into their calculations—with either formal models or with judgment—regarding the strength of the economy. The article, though it touches only on selected data relationships, should clarify how analysts carry information from one economic release into their view of the strength of other economic indicators.

REFERENCES

BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM. 1986. *Industrial Production: With a Description of the Methodology*. Washington, D.C.

GREENE, WILLIAM H. 1993. *Econometric Analysis*. 2d ed. Englewood Cliffs, N.J.: Prentice Hall.

HARRIS, ETHAN S. 1991. "Tracking the Economy with the Purchasing Managers' Index." Federal Reserve Bank of New York *Quarterly Review* 16 (Autumn): 61–69.

HARRIS, ETHAN S., AND CLARA VEGA. 1996. "What Do Chain Store Sales Tell Us about Consumer Spending?" Federal Reserve Bank of New York *Economic Policy Review* 2 (October): 15–35.

ROGERS, R. MARK. 1988. "Improving Monthly Models for Economic Indicators: The Example of an Improved CPI Model." Federal Reserve Bank of Atlanta *Economic Review* 73 (September/October): 34–50.

———. 1992. "Forecasting Industrial Production: Purchasing Managers' versus Production-Worker Hours Data." Federal Reserve Bank of Atlanta *Economic Review* 77 (January/February): 25–36.

———. 1994. *Handbook of Key Economic Indicators.* Burr Ridge, Ill.: Irwin Professional Publishing.

———. 1998. *Handbook of Key Economic Indicators.* 2d ed. Burr Ridge, Ill.: McGraw-Hill Professional Publishing.

U.S. DEPARTMENT OF COMMERCE. 1996. "Updated Summary Methodologies." *Survey of Current Business* (August): 81–103.