Economic Review

Federal Reserve Bank of San Francisco

Fall 1991 Number 4

Table of Contents

Opinions expressed in the Economic Review do not necessarily reflect the views of the management of the Federal Reserve Bank of San Francisco, or of the Board of Governors of the Federal Reserve System.

The Federal Reserve Bank of San Francisco's Economic Review is published quarterly by the Bank's Research Department under the supervision of Jack H. Beebe, Senior Vice President and Director of Research. The publication is edited by Judith Goff. Design, production, and distribution are handled by the Public Information Department, with the assistance of Karen Flamme and William Rosenthal.

For free copies of this and other Federal Reserve publicatons, write or phone the Public Information Department, Federal Reserve Bank of San Francisco, P.O. Box 7702, San Francisco, California 94120. Phone (415) 974-2163.

Have Large Banks Become Riskier? Recent Evidence from Option Markets

Mark E. Levonian

Senior Economist, Banking and Regional Studies, Federal Reserve Bank of San Francisco. I am grateful to Brian Cromwell, Randall Pozdena, and the editorial committee of Frederick Furlong, Brian Motley, and Jonathan Neuberger, for numerous helpful comments and suggestions.

This paper examines trends in risk at the largest U.S. commercial banks during the late 1980s. Prices of exchange-traded options on bank equity are used to derive several measures of banking risk. The results show that the riskiness of bank assets and activities did increase at large banks during the period. However, market capital-asset ratios generally rose, leaving the burden on the deposit insurance fund little changed. Hence, while the results support the notion that banks now engage in a riskier business than previously, the general increase in capital has been sufficient to hold overall banking risk relatively constant.

A series of events during the 1980s generated renewed concern about the condition of large commercial banking firms in the United States. Losses on loans to less-developed countries, energy-related loans, and problems related to real estate markets in various regions reduced the financial strength of many banks. At roughly the same time, large banks expanded their activities in a variety of nontraditional areas—including securities underwriting, trading of interest-rate and foreign-exchange-rate instruments, and financing of highly leveraged transactions—many of which appeared to hold potential for, and in some cases actually caused, significant losses. The combination of reductions in financial strength and expansion into new activities raised the fear that large U.S. banks became significantly riskier during the 1980s.

The riskiness of banks might be of less general interest if not for its impact on federal deposit insurance. The liability borne by the insurance fund roughly depends on expected losses due to bank failures. If banks became riskier during the 1980s, losses may have become more likely, and the liability of the insurance fund may have grown substantially. Large banks are of particular concern, not only because individually they are large components of the banking system, but also because they, more so than smaller banks, have been involved in the nontraditional activities mentioned above. If risk has increased at large banks, then some form of regulatory response—for example, increased bank capital standards, or restrictions on bank activities—may be desirable. Alternatively, the adverse conditions of the 1980s may not have affected risk materially, in which case calls for an active regulatory response to reduce risk are misdirected.

The well-publicized decline in the bank deposit insurance fund, from 1.19 percent of insured deposits at the beginning of 1985 to 0.70 percent at the end of 1989, may seem to constitute clear evidence of an increase in risk. However, Shaffer (1991) has demonstrated a fairly high probability of problems of this magnitude, without any change in the distribution of losses. That is, it is quite likely that the FDIC could experience such losses purely through a series of bad years, random "bad luck," without any change in banking risk. Thus, trends in banking risk cannot be examined simply by observing changes in the reserves of the insurance fund over time.

This paper focuses on the evolution of risk at nine of the largest U.S. bank holding companies. Changes in several measures of banking risk are examined. The primary contribution of the paper is the use of a new type of data: The prices of exchange-listed options on bank stocks. The size of the deposit insurance liability at any point in time depends critically on the prospects of insured banks. Option prices provide a unique source of information on market beliefs regarding both future risk and current financial condition, information that can be used to construct estimates of the risk to the insurance fund.[1] In addition, this paper provides empirical implementation of a relatively new model of insured banks, which is used to link the option price data with stock price and financial data to derive measures of banking risk.

The use of options data restricts the analysis to the second half of the 1980s, when options on most bank stocks began trading. Although some of the events that reduced the financial strength of banks—for example, losses on energy loans and loans to less-developed countries—occurred during the early 1980s, others, such as increased involvement in securities underwriting and trading of relatively exotic financial instruments, also had effects in the second half of the decade. A study by Furlong (1988) provides a useful complement to the present paper: Furlong applied similar methods to analyze changes in the first half of the 1980s.

The first section of this paper defines three measures of banking risk: The volatility of returns on bank asset portfolios, the size of bank capital cushions as reflected in capital ratios, and the overall liability imposed by banks on the deposit insurance fund. Section II presents a contingent-claim model of an insured bank and formalizes the three measures of banking risk. Section III describes a method for computing the market value and volatility of bank assets for use in the contingent-claim model; the section also discusses in detail the reasonable range of values for some key unobservable parameters of the model. Section IV explains the methodology and data used to infer estimates of risk from the market prices of bank stock options and presents the resulting estimates. Section V contains the major results of the paper regarding changes in risk at large banks in the late 1980s. The last section of the paper provides some concluding remarks.

I. Risk in Banking

From an ultimate policy perspective, probably the most important banking-related risk is the risk of losses to the deposit insurance fund. The expected value of these losses at any particular bank depends (1) on the probability that the bank's assets will fall short of its liabilities, thereby exhausting the bank's own capital, and (2) on the size of the shortfall if losses should occur. Of course, payouts from the insurance fund also depend on the degree of coverage offered by the insurer. But under any given coverage policy, the probability that a bank will fail and the size of the necessary insurance payment in the event of failure combine to determine the present value of the liability of the fund.

Both factors in turn reflect two broad types of banking risk. The first is financial risk, which depends on bank capital: The probability that a bank will fail varies inversely with the bank's capitalization for a given combination of bank assets and activities. Capital is defined in this paper as the difference between assets and liabilities, exclusive of deposit insurance. (Equity and capital are not identical in this context; the value of equity *includes* the value of the protection afforded by federal deposit insurance, which limits the liability of stockholders and protects depositors from losses due to unanticipated declines in the value of assets.) Capitalization is expressed most conveniently in terms of the capital ratio (the ratio of capital to total assets), with a higher capital ratio implying lower financial risk, all else equal.

The second broad type of risk is operating risk, which depends on the riskiness of the bank's asset portfolio. This risk is measured most directly by the variability of the rate of return on bank assets, referred to as the "volatility" of the bank's asset portfolio. Volatility is quantified by the statistical standard deviation of percentage changes in the value of bank assets. A bank with higher asset volatility is more likely to fail (and, if it fails, is more likely to impose a larger burden on the insurance fund) for any given capital ratio.

Calculating capital ratios and asset volatilities in order to measure risk is not a simple task. The relevant capital ratios must be computed from the market values of assets and liabilities, but market values often are not observable. Similarly, the relevant volatilities are the volatilities of the actual economic returns on the market value of assets; these returns may be very different from the observable accounting returns on the book value assets. Hence, to measure either financial risk or operating risk, or to combine the two into an estimate of the insurance fund liability, market values somehow must be calculated.

Although the market value of bank assets is not observable, the market value of bank equity *is* observable, since large banks have shares traded on stock exchanges. If the stock market is efficient, then the market value of equity reflects the market value of assets (although it also may depend on other factors, including the value of deposit insurance); hence, a model that correctly specifies the relationship between equity and asset values can be used to infer the latter from the former. In addition, the volatility of equity reflects the unobservable volatility of the underlying assets, again suggesting the possibility of inferring one from the other. The next section describes a model that, in addition to filtering out the effects of deposit insurance on equity values, relates the market value and volatility of bank equity to the market value and volatility of bank assets to permit inferences from observed market data.

II. Model of an Insured Bank

Merton (1974) applied contingent-claim techniques to the general problem of valuing the debt and equity of levered firms; in Merton (1977), the same techniques were applied specifically to insured banks. Following Merton's initial theoretical work, Marcus and Shaked (1984) implemented a similar model to derive empirical estimates of bank capital, asset volatility, and the size of the deposit insurance liability. In these models, banks have market value of assets A_t (excluding the value of deposit insurance), and total liabilities maturing with face value B_T at date T, at which time the bank is examined by regulators and is closed if assets do not equal or exceed liabilities. These assumptions imply a value of equity E at date T of:

$$(1) \qquad E_T = \begin{cases} A_T - B_T & \text{if } A_T \geq B_T \\ 0 & \text{if } A_T < B_T \end{cases} .$$

At any time prior to T, the total market value of a bank's equity is equal to the discounted value of this payoff structure. Equity in the model is a contingent claim (a positive payoff to equity is contingent upon the bank being solvent at T), and its value at any earlier point in time can be calculated using the same valuation techniques used in pricing other contingent claims, such as options.

Levonian (1991) revised this contingent-claim model of insured banks to incorporate both a flexible regulatory closure threshold and positive bank charter value.[2] The inclusion of charter value recognizes the fact that, in practice, banks operate under special charters granted by either state or federal authorities; because the supply of bank charters is limited, the positive value conferred by a charter is not competed away. Charter value is modeled as being a fraction ϕ of liabilities, and as being received by bank equity holders at date T only if the bank is not closed by regulators.[3]

In Merton (1977), banks are closed if they are insolvent at date T. However, in reality regulators have some discretion regarding closure, and the regulatory closure threshold need not be the point of actual insolvency. Banks may be closed while net worth is positive, or may be allowed to continue operating with negative net worth. If the regulatory rule is that a bank is closed if its capital ratio is less than c, then the value of equity at the monitoring date T is

$$(2) \qquad E_T = \begin{cases} A_T - B_T + \phi B_T & \text{if } k_T \geq c \\ 0 & \text{if } k_T < c \end{cases} ,$$

where c is not necessarily equal to zero, and the capital ratio k is defined as

$$(3) \qquad k_T \equiv \frac{A_T - B_T}{A_T} .$$

(A minor difference between this model and (1) is that the closure rule is stated in terms of the capital ratio rather than in terms of the relationship between assets and liabilities. Note that if $c=0$, then $k>c$ implies $A>B$.) Banks that remain open at date T experience a lump-sum increase in value from the rents conferred by a banking charter, where ϕB_T is the value of those rents.[4]

Realistically, E_T can never be negative, since the owners of a bank can always exercise their right of limited liability to walk away from a losing proposition. Thus it must be true that banks are closed at capital ratios above the level at which the charter value would be completely offset by negative net worth; that is, at the point $k=c$, it must also be true that $A-B-\phi B \geq 0$. Rewriting this restriction based on the definition of the capital ratio, the closure threshold must satisfy $c \geq -\phi/(1-\phi)$. If the closure threshold were set lower, regulators would be forced to inject funds—an outright gift, not just a loan—to induce some low-capital banks (those with $c<k<-\phi/(1-\phi)$) not to close voluntarily. The injection would have to be large enough to bring assets, and hence the capital ratio, back up to the minimum level of $-\phi/(1-\phi)$. While the FDIC does sometimes provide so-called open bank assistance, the actual extent of any wealth transfer is not obvious, since the emergency funding generally must be repaid by the surviving institution. In such cases, any net injection of capital comes in the form of FDIC acceptance of a below-market rate on the funds. As an alternative and less complex treatment of this possibility, any assistance anticipated by the market is assumed to be capitalized into ϕ, and c is always no less than $-\phi/(1-\phi)$.

As in most applications of contingent-claim methods, assets are assumed to follow a stochastic process given by

$$(4) \qquad dA = \mu_A A(t)dt + \sigma_A A(t)dz,$$

where μ_A is the expected instantaneous periodic rate of return on assets, t is a time index, dz is the differential of a Wiener process, and σ_A is the instantaneous standard deviation of the rate of return on assets, or asset volatility. Let the date $t=0$ represent the present, and let unsubscripted variables denote present values. Using standard methods for valuing contingent claims (see Smith 1976), the present value of equity with date T payoff as given in (2) is

$$(5) \qquad E = AN(x) - BN(x - \sigma_A\sqrt{T}) + \phi BN(x - \sigma_A\sqrt{T}),$$

where

$$(6) \qquad x \equiv \frac{\ln\left(\frac{(1-c)A}{B}\right) + \frac{\sigma_A^2 T}{2}}{\sigma_A\sqrt{T}}$$

and $N(\bullet)$is the cumulative standard normal distribution function. Equity is essentially a call option on assets, plus an additional lump sum equal to the expected present value of the charter.[5] The first two terms in (5) represent the familiar option value; the third term is the charter value ϕB weighted by a factor that is closely related to the probability that the bank will remain open.[6]

Measures of Banking Risk

Given this theoretical framework, the central issue of this paper can be posed more explicitly. In particular, financial risk has increased at large banks if the market value capital ratio k defined in (3) has decreased; operating risk has increased at large banks if the volatility of assets σ_A has increased.

The deposit insurance liability, which combines the effects of both types of risk, also can be calculated explicitly once values for A and σ_A have been obtained. The deposit insurance contract is another contingent claim and can be evaluated using the same methods. All of the banks in the sample are sufficiently large that the market has good reason to expect that all creditors will be protected from losses in the event of a failure; hence, the contingent deposit insurance liability should be modeled under the assumption that the claim covered by the insurer is B, even though not all liabilities are formally insured. The typical method of resolution when such a large bank fails is to locate a purchaser for the failed institution; the acquirer receives the assets and the charter of the failed bank and assumes all of the liabilities. If the liabilities assumed by the acquirer exceed the value of the assets and charter, the deposit insurer makes up the difference. Thus, the insurance fund pays the acquirer $B-(\phi B + A) = (1-\phi)B - A$ if that difference is positive, and otherwise pays nothing. Formally, the insurer's payout is

$$(7) \qquad V_T = \begin{cases} (1-\phi)B_T - A_T & \text{if } B_T > A_T + \phi B_T \\ 0 & \text{if } B_T \leq A_T + \phi B_T \end{cases}.$$

Again using standard contingent-claim valuation techniques, the value of the contingent payout in (7) is

$$(8) \qquad V = (1-\phi)BN(y + \sigma_A\sqrt{T}) - Ae^{-\gamma T}N(y),$$

where

$$(9) \qquad y \equiv \frac{\ln\left(\frac{(1-\phi)B}{Ae^{-\gamma T}}\right) - \frac{\sigma_A^2 T}{2}}{\sigma_A\sqrt{T}}$$

The dividend rate, γ, appears in (8) and (9) because dividend payments directly reduce the assets available to the deposit insurer in the event of failure. (Note that γ is the rate of dividend payments relative to assets, not equity.) The overall risk to the deposit insurance fund posed by large banks has increased if V has increased.

III. A Method for Computing Asset Values and Asset Volatility

Assuming that the value of bank equity is determined as in (5) it is possible to work backward from the stock market prices of large publicly traded banks to infer the market value of assets and asset volatility. Various realistic values can be assumed for bank liabilities B, the regulatory monitoring interval T, the capital ratio closure threshold c, and the charter value ratio ϕ. The two remaining unknowns in (5) are the value of assets σ_A and the volatility of assets σ_A.

Obviously, a single equation cannot be solved for two unknowns; a second independent equation is needed. Merton (1974) suggests applying Itô's Lemma to the expression for the value of equity, to yield a second equation relating the volatility of equity and the volatility of assets. Merton derives the relationship[7]

$$(10) \qquad \sigma_E = \sigma_A \frac{\partial E}{\partial A} \frac{A}{E}$$

An intuitive grasp of (10) follows from considering the case in which bank stockholders do not have limited liability for the debts of the bank. In that case, the contingent aspect that makes equity behave like a call option on assets disappears, and the value of equity changes one-for-one with the value of assets.[8] Then $\partial E/\partial A = 1$, and (10) reduces to $\sigma_E = \sigma_A(A/E)$, with the straightforward interpretation that the volatility of equity is simply the "levered-up" volatility of the underlying assets. However, with limited corporate liability, equity becomes somewhat less sensitive to changes in asset values, as gains and losses are shared partially with debtholders. Then $\partial E/\partial A < 1$, and σ_E falls relative to σ_A.

In the present case, differentiation of (5) yields

$$(11) \qquad \frac{\partial E}{\partial A} = N(x) + \frac{\Theta B N'(x - \sigma_A\sqrt{T})}{A\sigma_A\sqrt{T}},$$

where $N'(\bullet)$ is the standard normal density function and $\Theta = 1/(1-c) - (1-\phi)$. Using (11), the expression in (10) can be rewritten as

$$(12) \qquad \sigma_E = \frac{AN(x)\sigma_A\sqrt{T} + \Theta BN'(x - \sigma_A\sqrt{T})}{E\sqrt{T}}.$$

Equation (12) depends on all of the same variables as equation (5). If σ_E is observable, then under identical assumptions regarding the parameters of the model, this equation also has A and σ_A as the only unknowns, and (5) and (12) can be solved simultaneously for values of the two unknown variables.[9]

As noted above, solving these two equations to obtain market assets and volatility requires making assumptions about the other parameters: bank liabilities B, the regulatory monitoring interval T, the capital ratio closure threshold c, and the charter value ratio ϕ. The market value of bank liabilities is assumed to be approximately equal to book value, since the bulk of bank liabilities are short term. The monitoring interval is assumed to be one year; this corresponds roughly to bank examination frequency. Assumptions regarding the charter value and closure threshold assumptions require more detailed explanation.

The Charter Value Ratio

Previous approaches to estimating bank charter value are inappropriate for this analysis. For example, Keeley (1990) divides the sum of book value liabilities and market value equity by book value of assets, and uses this ratio as a proxy for Tobin's q to examine changes in charter value; Kwan (1991) applies a similar approach based on q. Such estimates based on the market value of bank equity cannot capture the concept of charter value as defined in this paper, because they do not separate the effect of deposit insurance from other components of measured charter value. Some other method must be used to define a reasonable value for ϕ.

In practice, the value of a bank charter is likely to manifest itself in nonmarket interest rate spreads: either a rate of return on bank loans in excess of the required rate for that level of risk, or a below-market rate of interest on deposits, or some combination of the two. Hence, information on deposit and loan spreads can be used to develop an estimate of ϕ.

On the deposit side, if the bank charter gives the bank the ability to set $r_d < r_f$ and still attract federally insured deposits, the basic contingent claim model of bank equity must be modified; without the lump-sum charter value, equation (5) becomes

$$(5') \qquad E = AN(x') - Be^{(r_d - r_f)}N(x' - \sigma),$$

where T = 1 without loss of generality and

$$(6') \qquad x' = \frac{\ln(A/B) + (r_f - r_d + \sigma^2/2)}{\sigma}.$$

This can be viewed as an alternative formulation of the basic model presented in (5), in which the charter value is received as a flow over time in the form of a rate spread rather than as a lump sum ϕB at the end of the period. Comparing the two forms of the model, if the relatively small effect of $(r_d - r_f)$ on x' is ignored, then $(1-\phi)=e^{(r_d-r_f)}$. In this case, the charter value ratio can be approximated by the deposit interest rate spread, $\phi \approx r_f - r_d$, since for realistic spreads it will be true that $e^{(r_d-r_f)} \approx 1-(r_f-r_d)$.

If instead the charter allows the bank to earn an above-market rate of return on assets, a variant of the contingent claim model is appropriate. McDonald and Siegel (1984) consider the case of a contingent claim on an asset earning a rate of return different from the appropriate risk-adjusted rate. If Δ is the loan spread—the rate of return on loans held by the bank minus the required rate of return for assets of comparable risk—then the value of bank equity can be expressed as

$$(5'') \qquad E = Ae^{\Delta}N(x'') - BN(x''-\sigma),$$

where again T = 1 and

$$(6'') \qquad x'' = \frac{\ln(A/B)+(\Delta+\sigma^2/2)}{\sigma}.$$

This version of the model, with A multiplied by a factor e^{Δ} which is positively related to charter value, suggests that the charter value should be modeled as being proportional to assets rather than liabilities in this case. However, recognizing that $A=B/(1-k)$, that for most banks $N(x'') \approx N(x''-\sigma)$, and that $e^{\Delta} \approx 1+\Delta$ for realistic values of Δ, and ignoring the trivial effect of Δ on x'', a value of Δ greater than zero increases the value of bank equity by an amount equal to

$$(13) \qquad \Delta AN(x'') \approx \frac{\Delta B}{1-k}N(x''-\sigma).$$

Thus, in the context of the model presented above in (5), the effect on bank equity is roughly equivalent to setting $(1-\phi) = -\Delta/(1-k)$.

If a chartered bank has positive spreads on both the deposit and the loan side of the business, the joint effect can be approximated as

$$(14) \qquad (1-\phi) \approx 1-(r_f-r_d)-\frac{\Delta}{1-k},$$

where

$$(15) \qquad \phi \approx \frac{\Delta}{1-k} + (r_f-r_d).$$

The approximation in (15) was combined with data on interest rates to provide a sense of reasonable values for ϕ. The deposit rate spread was proxied by the difference between the rate on six-month certificates of deposit (the national average from *Bank Rate Monitor*) and the secondary market yield on six-month U.S. Treasury bills. The loan spread was proxied by the difference between the weighted average interest rate on short-term commercial and industrial bank loans (from the Federal Reserve's survey of terms of bank lending) and the rate on one-month commercial paper. (One-month commercial paper was used because it was closest to the average maturity of bank loans reported in the terms-of-lending survey.) Combining the average values of the interest rate spreads (based on quarterly data for the sample period) with values of k between 0 and 10 percent produced estimates of the value of ϕ in the relatively narrow range of 0.016 to 0.018. Since substantial approximation error is likely, these estimates should be taken only as indicative of the neighborhood of the charter value ratio; values for ϕ of 0.01 and 0.02 are used in Section V to bracket a reasonable range.

In the model, the charter value ratio is assumed to be constant over time. In reality, interest rate spreads fluctuate, and a systematic trend might cause calculations based on the assumption of constant ϕ to be biased. To test for the existence of a trend, the sum of the interest rate spreads was regressed on a time variable, with a correction for first order autocorrelation. The trend coefficient was positive but insignificant (at the 5 percent level) during the sample period. Thus, the assumption that ϕ is constant over time probably is innocuous.

The charter value ratio also is assumed to be identical for all banks in the sample. It is possible that differences in managerial ability, location, and other factors might cause different banks to reap different benefits from their charters. However, the nine banks in this sample are sufficiently similar in size and character that interbank differences in ϕ are unlikely to be major, despite some variation in business strategy among the sample firms.

The Closure Threshold

It is unlikely that regulators would seize any of the large banks in the sample at positive market value capital ratios. Hence, the assumed value of c almost certainly should be zero or less. However, as noted in Section II, a credible closure point cannot be so low that the charter value is completely exhausted before the bank is closed; that is the threshold must satisfy $c \geq -\phi/(1-\phi)$. Since the charter value ratio is assumed to be in the range of 0.01 to 0.02, the closure threshold cannot be less than about 0.01 if ϕ is on the low side at 0.01, or 0.02 if ϕ is 0.02.

As with the charter value ratio, there is some possibility that c varies either over time, or across banks in the sample, or both. Substantial interbank variation within this sample seems unlikely, for the same reasons given above in the discussion of ϕ. As for variation over time, if closure policy changed during the period, there should be some evidence of a change in the loss experience of the deposit insurance fund. To examine this possibility, FDIC losses resulting from bank closures (deposit payoffs, deposit transfers, and purchase and assistance transactions) were computed from the FDIC's *Annual Report* and divided by the deposits of closed banks to construct a loss ratio. This ratio was roughly the same at the end of the sample period as at the beginning, and a regression of the ratio on a time variable revealed no significant trend during the 1980s. Thus, the assumption of constant c probably is a reasonable approximation.[10]

IV. Equity Volatility From Traded Option Prices

The two-equation approach to deriving estimates of A and σ_A for a sample of banks has been used previously by Marcus and Shaked (1984), Ronn and Verma (1986), Furlong (1988), and Kendall and Levonian (1991), among others. All of these earlier studies used the standard deviation of historically observed stock returns for equity volatility σ_E. But conceptually, the relevant volatility is the *expected* volatility over the period from $t=0$ to $t=T$. Use of historical volatility assumes that expectations at each point in time are formed adaptively, and therefore reflect realizations over some recent interval. If traders form their expectations of σ_E using information in addition to historical returns, then the historical standard deviation may be a poor proxy for the expected volatility required in the contingent-claim framework.

A more direct measure of expected volatility is both desirable and available. Options on bank stocks trade on several U.S. options exchanges; the prices of stock options are known to depend in part on the expected volatility of the underlying stocks. Using an option pricing model, values of expected σ_E can be inferred from traded option prices.[11] Because exchange-traded stock options in the United States have American terms (meaning that the holder may choose to exercise prior to expiration), expected volatilities are inferred from an American option pricing model developed by Barone-Adesi and Whaley (1987) given by

$$(16) \qquad C = Se^{-\delta\tau}N(z) - Xe^{-r\tau}N(z-\sigma_E\sqrt{\tau}) + P(S,X,\tau,\sigma_E,r,\delta),$$

where C is the value of a τ-period American call option with exercise price X on a stock with price S and continuous dividend rate δ (dividends relative to equity), r is the risk-free interest rate, and z is defined as

$$(17) \qquad z \equiv \frac{\ln\left(\frac{S}{X}\right) + \tau(r-\delta) + \frac{\tau\sigma_E^2}{2}}{\sigma_E\sqrt{T}}$$

The first two terms in (16) give the value of a European option, which cannot be exercised prior to the expiration date. The function $P(S, X, \tau, \sigma_E, r, \delta)$ is an approximation of the early exercise premium (the difference in value between an American option and a European option due to the possibility of early exercise), the exact form of which is derived by Barone-Adesi and Whaley. They demonstrate that their approximation works well for the range of expirations used in this paper.

All of the variables in (16) are observable in the financial press or elsewhere, with the exception of σ_E. Although (16) cannot be inverted for σ_E, standard numerical techniques can be used to find the unique value of σ_E satisfying (16), which is referred to as the "implied volatility" of the stock. This implied volatility can then be used in (12) to solve for σ_A and A.

Levonian (1988) showed that implied volatilities of bank stocks differ substantially from volatilities calculated using historical stock returns. Both Black and Scholes (1972) and Latané and Rendleman (1976) used tests based on observed option prices to show that historical volatility is inferior to implied volatility as a predictor of future volatility; Schmalensee and Trippi (1978) obtained similar results. Hence, volatilities implied by option prices should provide better information about the riskiness of bank stocks, and consequently about the various types of banking risk, than would volatility estimates based on historical stock returns.

However, the use of implied volatility is not without cost. Since far fewer banks have traded options than have traded stock, the sample size is reduced substantially. As always in empirical research, small sample size may bias the results. Thus, it is possible that the results may fail to represent adequately the riskiness of the banking industry as a whole, even though the volatility estimates for each individual bank are likely to be superior.

Options Sample

The sample for this paper consists of listed call options from the various options exchanges in the United States, sampled at the ends of the second and fourth quarters of the five years 1985 through 1989, for nine large banking firms: BankAmerica, Bankers Trust, Chase Manhattan, Chemical, Citicorp, First Chicago, J.P.Morgan, Manufacturers

Hanover, and Security Pacific. This group comprised nine of the top ten U.S. banks at the beginning of 1985, ranked by assets. (The tenth, First Interstate, also has exchange-listed options, but trading was too infrequent to allow construction of a reliable time series.) These are options on the common stocks of bank holding companies, not banks; however, to the extent that the predominant assets of holding companies are banking-related, the implied volatilities provide information on bank asset risk. The term "bank" is used to refer to these firms throughout the paper.

The interest rates used in the option pricing model were the yields-to-maturity on the U.S. Treasury bills maturing closest to the expiration date of each option. The dividend rates were computed by assuming that expected dividends during the life of each option were identical to dividends actually paid and then calculating equivalent continuous rates.[12] Last-trade-of-the-week call option prices and stock prices for the nine banks were collected from published listings in *The Wall Street Journal*.

Day and Lewis (1988) describe two sources of potential bias in the use of published prices. One is the problem of nonsynchronous trading, that the last option trade for any given bank may not have occurred at the last observed stock price; the option valuation model in (16) requires the use of a contemporaneous stock price. The other is that stocks and options trade with a spread between the bid price and the ask price, and reported trades may occur at either the bid or the ask or at prices in between, making it impossible to observe a precise estimate of value. Day and Lewis argue that estimates of implied volatility should incorporate information from the prices of several different options on the same stock in order to minimize the potential bias. Studies of option volatility have used a variety of methods for combining different options on a single underlying security (for example, compare Day and Lewis to Latané and Rendleman or Schmalensee and Trippi.

To deal with these problems, for each bank only the two options with exercise prices closest to the current stock price are used. That is, for all of the options with exercise prices above the current stock price, the option with the lowest exercise price is selected; in addition, for all of the options with exercise prices below the stock price, only the one with the highest exercise price is selected, for each bank. Day and Lewis show that trading volume is concentrated in these "near-the-money" options; for the index options they examine, about 70 to 90 percent of the volume is in options with exercise prices just above and just below the current stock price. Even a cursory review of published options sales data confirms that this relationship is true in general. As Day and Lewis point out, any lack of synchronization between the closing stock price and the closing option price will be minimized for these options. They also note that the percentage bid-ask spread is less for these options, reducing the second source of bias as well. (Feinstein 1988 provides a discussion of other desirable properties of near-the-money options for the purpose of inferring volatilities from prices.)

It is possible that at any point in time market traders anticipate that volatility will change in some predictable way over time. In that case, the options from which the stock volatility is inferred should have expirations identical to the regulatory monitoring interval for banks, assumed to be one year in this paper. However, until very recently exchange-traded stock options were restricted to expirations of less than a year; moreover, the most active trading generally occurs in options with short expirations. Thus, short-term options are likely to yield superior estimates of expected σ_E, and these estimates can be used in the bank equity model provided that volatility is not expected to change drastically between the expiration date of the option and the end of the regulatory monitoring period. However, using the shortest expirations may introduce other problems; Day and Lewis document a statistically significant increase in implied volatility for options as the expiration date approaches, especially in the last few trading days. They attribute this effect to technical factors related to the unwinding of hedged positions. To achieve a balance between both types of distortions, the options sample for this study consists of the shortest-term options for each bank, but with a minimum time to expiration of one month. This sample selection process is similar in spirit to that used by Schmalensee and Trippi (1978).

An additional complicating factor is that unusual events or changes in option market liquidity might cause option prices from a single week to be unrepresentative of the true riskiness of banks. To minimize this problem, prices were sampled for three consecutive weeks surrounding each semiannual sample point: The week of the financial reporting date, one week before that date, and one week after. For each bank in each of the weeks, prices of the two call options with exercise prices nearest to the underlying stock price and with shortest time to expiration (but exceeding one month) were collected. Implied stock volatility was calculated for each option, and all six options averaged for each bank at each semiannual date. The procedure produced 90 estimates of volatility (ten semiannual observations for nine banks).[13]

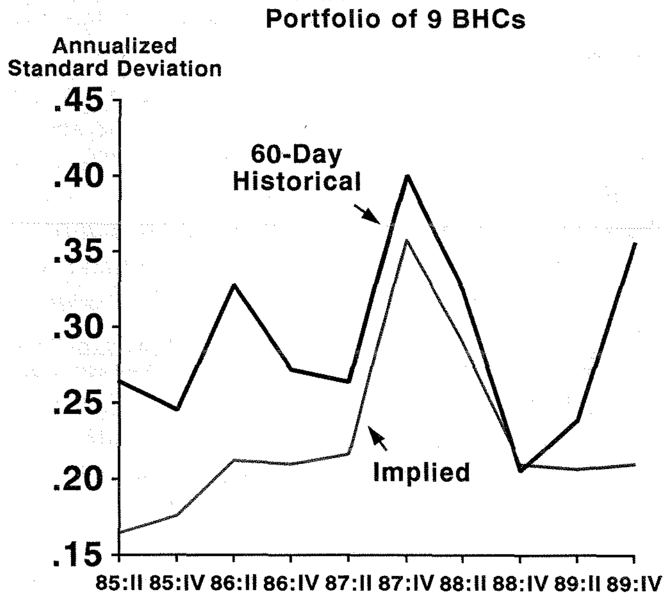
Stock Volatility Estimates

A weighted average of the nine banks was computed for each time period to summarize the resulting implied stock volatilities and to present the pattern of changes during the late 1980s. The weights for the observations were the market value of each bank's equity (stock price multiplied by number of shares) divided by the total market value of equity of all nine banks for that date. The resulting weighted average can be viewed as an index of implied bank stock volatility, with heavier emphasis given to banks that are larger components of total bank stock market value. (Alternatively, if returns on bank stocks were perfectly correlated, this average would equal the volatility of a stock portfolio consisting of equal percentages of each bank's equity, for example 5 percent of Bank A, 5 percent of Bank B, etc.) The results are presented in Chart 1, with volatility stated in annual terms (that is, the figures can be interpreted as standard deviations of annual percentage changes in the prices of the stocks). A similarly constructed index of historical volatility is presented for comparison. Historical volatility was computed as the annualized standard deviation of stock returns for the 60 trading days (roughly three months) preceding the end of the quarter.

One notable feature of Chart 1 is the upward spike in both implied and historical volatility in the fourth quarter of 1987. Implied volatility rose from 22 percent to 36 percent, and historical from 26 percent to 40 percent. This spike corresponds to the period immediately following the stock market crash of October 1987. Over the following year, volatility returned to levels similar to those preceding the crash. This pattern was not unique to bank stocks: Schwert (1990) documents similar effects for implied and historical volatility for the overall stock market as measured by the S&P500 stock index.

While the patterns of historical and implied volatility around the time of the crash are similar, there are noticeable differences in the rest of the sample period. Historical is almost always higher than implied for this sample, and three of the nine quarter-to-quarter changes are opposite in sign. The fourth quarter of 1989, in which the divergence is especially pronounced, provides an excellent example of the primary drawback of historical volatility. The high standard deviation of realized returns for the fourth quarter of 1989 is due to stock price movements on two dates,

Chart 1
Stock Return Volatility



Friday October 13 and Monday October 16. On Friday, the failure of UAL Corp. to obtain financing for a leveraged buyout precipitated a large decline in the overall stock market. Stocks of large banks were hit especially hard, apparently because the news was taken as a signal of a fundamental change in a major line of business. (Citicorp had the largest percentage single-day drop at 16 percent, and J.P. Morgan had the smallest at 5.5 percent.) On Monday, stock prices increased, recovering a portion of the value lost on the preceding Friday. A 60-day historical volatility calculation treats returns from these days equally with the other 58 days in the period. In reality, it is likely that traders viewed these two days as extreme events, and that by the end of December traders gave them little weight in formulating expectations of bank stock volatilities. The lower implied volatilities from option markets for 89:IV are direct reflections of expectations at that date, automatically discounting any information that is irrelevant to future returns.

V. Changes in Banking Risk, 1985-1989

The stock volatility results hold some intrinsic interest, and are roughly comparable to the bank stock volatility results presented by Jonathan Neuberger in another article in this *Review*. (Neuberger examines changes in bank stock risk during the 1980s in greater detail, and investigates the relationship between bank stock returns and returns in the bond market and the overall stock market.) However, the main purpose of the preceding stock volatility computations is to provide the raw material for other calculations related to banking risk. In this section, estimates of the three measures of banking risk are presented for the second half of the 1980s. Asset volatilities are obtained from the simultaneous solution of equations (5) and (12), using the implied σ_E for each bank in each period as the input to (12).

The solutions for market asset values from (5) and (12) are used to compute market value capital ratios from equation (3). Finally, the liability of the insurance fund is calculated from (8), also using the two-equation solution values of σ_A and A.

As discussed in Section III, the market value of bank liabilities is assumed to be equal to book value, and the monitoring interval is set equal to one year. The actual dividends paid by each bank during each year are used to compute the dividend rate γ in (8), under an assumption that dividends were paid as expected. The earlier discussion of realistic ranges for the charter value ratio and the closure threshold concluded that $\phi=0.01$ and $\phi=0.02$ would provide a reasonable bracket for charter value, and that c should be less than or equal to zero, but no larger in absolute value than ϕ. Thus, four cases are considered for combinations of these two parameters:

C1: $c=0.00$, $\phi=0.01$ — Charters have low value, and banks are closed when they are insolvent in market value, that is, when $A<B$.

C2: $c=0.01$, $\phi=0.01$ — Charters have low value, and banks are closed when charter value is exhausted.

C3: $c=0.01$, $\phi=0.02$ — Charters have high value, and banks are closed before charter value is exhausted.

C4: $c=0.02$, $\phi=0.02$ — Charters have high value, and banks are closed when charter value is exhausted.

Comparing the results from the four cases provides insight into the sensitivity to changes in the assumptions. Three pairwise comparisons are most interesting:

C1 vs. C2: impact of a lower closure threshold when charter value is low.

C3 vs. C4: impact of a lower closure threshold when charter value is high.

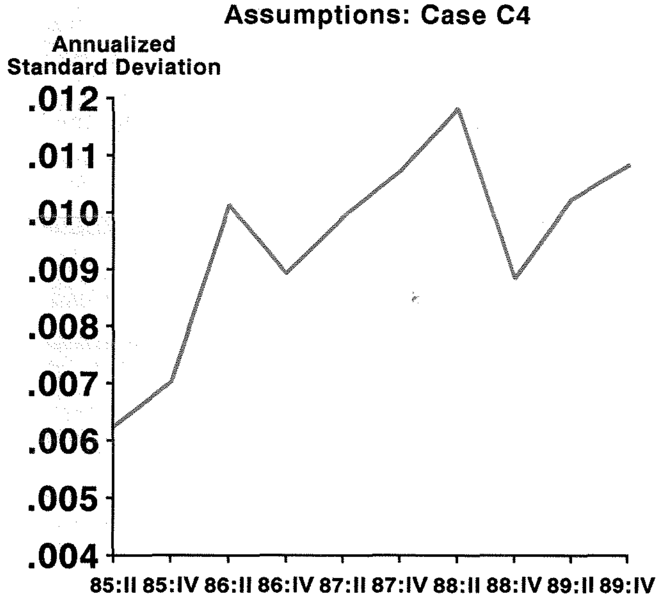
C2 vs. C3: impact of higher charter value with a fixed closure threshold.

Operating Risk

Weighted averages are constructed to summarize the individual bank results for each time period, with each bank weighted by liabilities relative to total liabilities of the nine banks for that date. The individual banks could instead have been weighted by equity (as with the stock volatilities above) or by assets. Other weightings produced very similar results.

Chart 2 shows the evolution of bank operating risk as reflected in annualized figures for asset volatility. Only case C4 is presented, because alternative assumptions about c and ϕ had very little effect on the estimates of σ_A. The general upward trend shows that operating risk did indeed increase at large banks during the late 1980s: The level of asset volatility in 89:IV was about 80 percent higher than in 85:II. At least in part, this increase may have been due to the expanded range of activities conducted by banks. The rate of increase in operating risk is consistent with that found by Furlong (1988) for the early 1980s, implying that bank assets became progressively more volatile across the entire decade. Interestingly, the increase in average asset volatility in the fourth quarter of 1987 was not dramatically different from adjacent periods. Hence, the significant jump in stock volatility in that period was *not* due to any great increase in the riskiness of bank assets as perceived by the market.

Chart 2
Asset Volatility

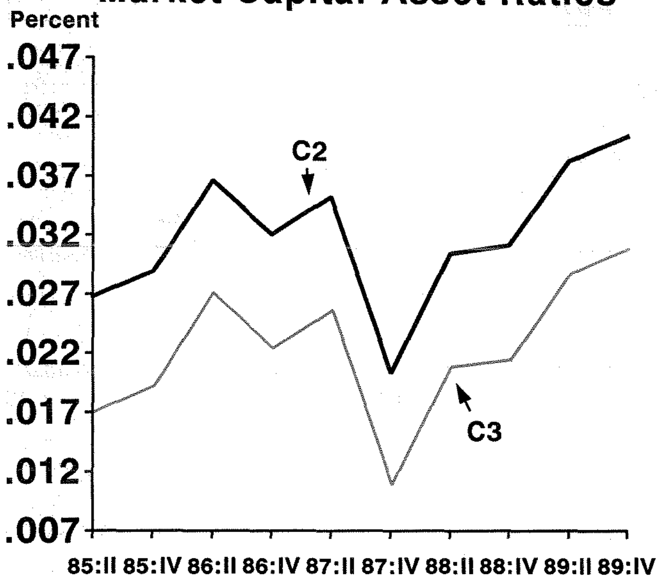


The pattern of asset volatility during the period is most heavily influenced by the larger banks in the sample, such as Citicorp, because of the use of a weighted average. The patterns for individual banks differ somewhat, although for each of the nine banks asset volatility was higher at the end of the sample period than at the beginning.

Financial Risk

Chart 3 shows weighted-average market capital ratio results for two illustrative cases, C2 and C3. Capital ratios were higher at the end of the sample period than at the beginning, implying that financial risk decreased at large

Chart 3
Market Capital-Asset Ratios



banks. (The pattern also is representative of the banks individually, except that several banks show slight declines in capital ratios in the fourth quarter of 1989.) However, market capital ratios suffered a tremendous hit in the stock market crash, from which they only gradually recovered over the next one-and-a-half to two years.

In the two cases omitted from Chart 3, C1 was essentially the same as C2, and C4 was little different from C3, demonstrating that variations in the assumed closure threshold have a trivial impact on the resulting capital ratio estimates. The only major distinction among the four alternative sets of assumptions was that variations in the assumed charter value produced capital ratios that differed by roughly the magnitude of the difference in the charter value ratio. The reason is straightforward. Recall that the capital ratio k is based on assets exclusive of the charter value; since a higher charter value should make bank stock more valuable all else equal, a *given* market value of equity can only be consistent with a *lower* market value of assets, and hence a lower market value of bank capital. Except for this charter value difference, the pattern over time is very similar in all four cases.

The decline in the market capitalization of banks resulting from the October 1987 crash evident in Chart 3 explains the pattern of implied stock volatility observed in Chart 1. Although Chart 2 showed that the riskiness of bank assets did not increase, the market value of equity fell as stock prices collapsed, and large banks suddenly became much more highly levered. The higher volatility of bank stock returns reflected in call option prices stemmed from traders' recognition that any given percentage change in the return on bank assets would translate into a much larger percentage change in return on the stock, when viewed relative to the lower base value of equity. (In terms of equation (10), the ratio A/E increased; for a given σ_A, this leverage increase caused σ_E to rise.) An interesting implication of the results presented here is that, at least for banks, the increase in volatility was of roughly the magnitude that should be expected after a decline in market equity-to-asset ratios of the extent experienced in the October crash.

The Deposit Insurance Liability

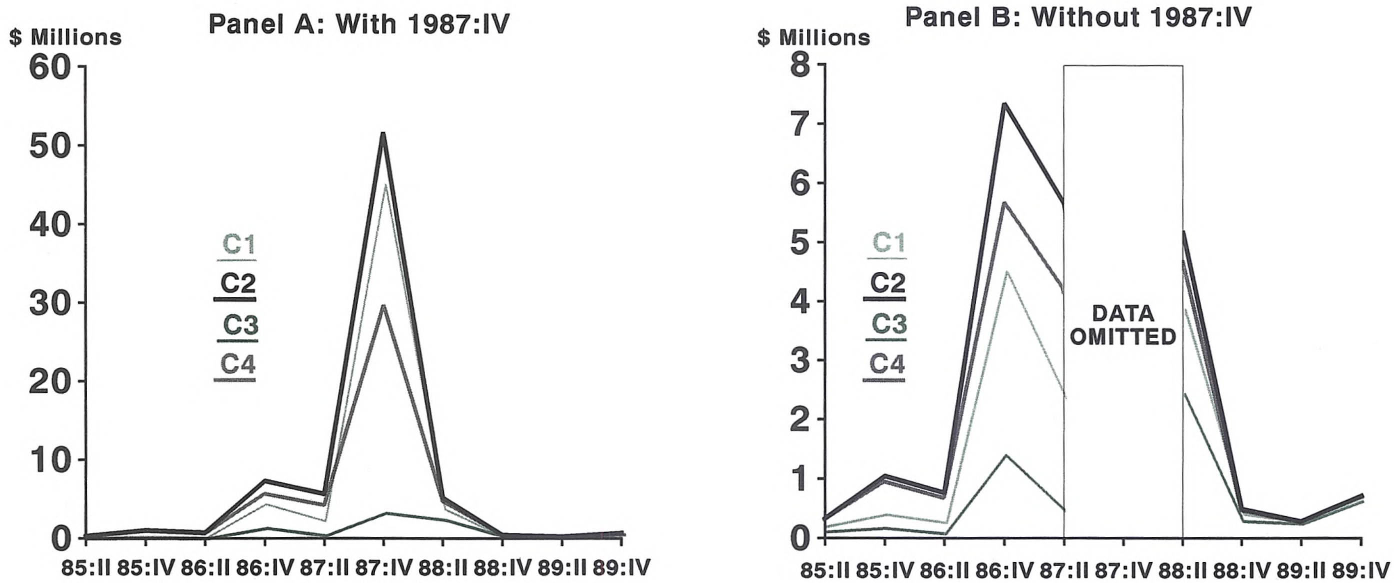
Chart 4 shows the path of the estimated deposit insurance liability over the sample period. Because the overwhelming impact of the stock market crash makes it hard to see the pattern of changes in the deposit insurance liability for other periods, Panel B displays the same results as Panel A, but with the 87:IV data omitted.

Panel A displays the total value of the liability at each date, summed across banks, for each of the four parameter cases. The stock market crash had a temporary but dramatic impact on the computed liability of the insurance fund. In case C1, for example, the liability rises from \$2.4 million in 87:II to \$45.3 million in 87:IV, then falls back to near the year-earlier level by 88:II. The reason is clear from Chart 3: The fall in market capital ratios associated with the decline in bank stock prices caused a large increase in financial risk at these banks, and the ensuing recovery in market value reversed the change. Each bank in the sample exhibits roughly the same pattern.

In all four cases in Panel B, the liability is only slightly higher at the end of the sample period than at the beginning. This conclusion would hold even if the correct values of c and ϕ changed during the period. For example, if regulators began allowing more poorly capitalized banks to remain open as time passed, so that C3 was most realistic at first and C4 was most realistic at the end, the 89:IV liability in case C4 is still little greater than the 85:II liability in case C3.

However, it is also true that the deposit insurance liability increased substantially during the middle part of the sample period, and that the increase predated the stock market crash. The change in risk was driven by the general increase in financial risk that began in 86:II, during a period of rising operating risk. The increase in market capital ratios toward the end of the sample period (assisted by the relatively large drop in asset volatility in 88:II)

Chart 4
Deposit Insurance Liability



brought the deposit insurance liability down to near its earlier level.[14]

Comparison of the various cases reveals that assumptions about the closure threshold and the charter value ratio have some impact on the computed deposit insurance liability at each point in time, although not a major impact. Comparing C1 to C2 and C3 to C4 shows that an earlier closure assumption reduces the total liability. The magnitude of the effect is roughly the same under either charter value assumption. (Note that this result does *not* demonstrate that instituting a policy of earlier closure would reduce risk to the deposit insurance fund. These derived results are conditional on the observed prices of bank stocks; an explicit policy change probably would generate a behavioral response by banks, and would alter the market value of bank equity, thus complicating any evaluation of a change in regulatory policy. This point is similar to the "Lucas critique" in macroeconomics.)

Comparing C2 to C3, different assumptions about the value of bank charters have a somewhat larger impact on the computed liability, for a given closure threshold assumption: The higher value of ϕ implies a smaller deposit insurance liability. This effect stems from the assumption that the deposit insurer uses the charters of failed banks to offset at least partially any required transfer from the insurance fund. It is interesting also to compare C2 to C4 in this regard. In both cases, the closure threshold is set so that the bank is closed as late as possible, when the charter value is completely exhausted by losses on bank assets. Even with this extreme closure assumption, higher charter value reduces risk to the insurance fund.

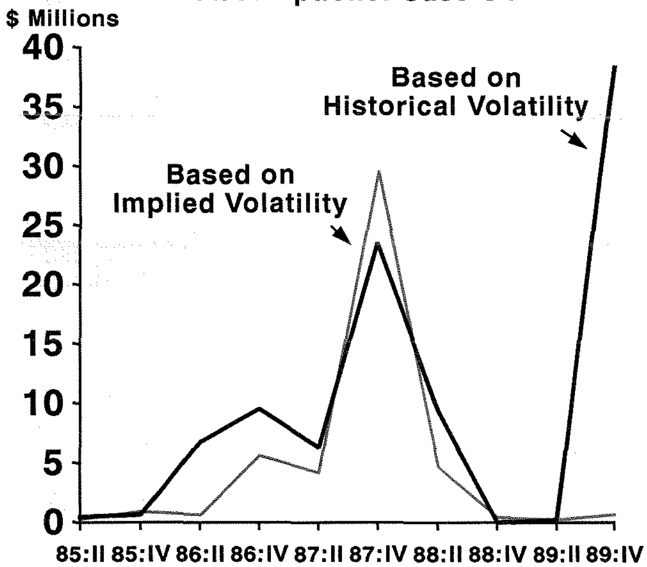
The inverse relationship between assumed charter value and the deposit insurance liability might seem to be at odds with the conclusion from Chart 3, which showed that higher assumed charter value increases financial risk. The apparent conflict can be resolved by recognizing that while higher charter values reduce measured capital ratios, the additional value is captured by the insurer in the event of a bank seizure, leaving little net effect on the insurance liability from changes in this parameter.

The moderate sensitivity of the deposit insurance results to alternative assumptions about the unobservable parameters makes it inappropriate to attach great weight to the specific dollar amounts of the liability; it is the general secular trends that are the important features for the aims of this paper. The overall pattern is clear, using any of the four parameter combinations: Despite a large increase in risk to the deposit insurance fund during the middle of the sample period, risk from these nine large banks was back down to a relatively low level by the end of the decade.

It is interesting to return to the comparison of historical stock volatility and implied stock volatility, and examine whether the difference between the two would affect the conclusions in this section. Chart 5 compares estimates of the total deposit insurance liability based on implied and historical σ_E, using C4 assumptions. The difference in the fourth quarter of 1989 is huge. Using historical volatility, the risk to the deposit insurance fund appears to be much

Chart 5
Effects of Implied and Historical Volatility on Deposit Insurance Liability

Assumptions: Case C4



higher at the end of the sample period than at the beginning. However, as noted in Section IV, the difference between historical and implied volatility in 89:IV is due almost entirely to stock price movements on two consecutive trading days, probably related to the collapse of the UAL leveraged buyout. In this case, implied and historical volatility lead to significantly different conclusions, and the results based on historical clearly are questionable.

Summary

The results in this section indicate that there was no significant overall increase in the riskiness of large banks during the late 1980s. The riskiness of bank assets and activities did increase at large banks over the five-year period studied, but concurrent with this increase in operating risk, financial risk fell as market capital ratios rose. The increase in market capital ratios was sufficient to prevent a large secular rise in the burden on the insurer. By the end of the sample period, these large banks posed little more risk to the deposit insurance fund than at the beginning of the period.

It is not surprising that capital increased as asset volatility rose, since the regulatory guidelines in effect during this period explicitly required banks engaged in riskier activities to maintain higher capital ratios (Board of Governors 1985). The positive relationship is evident in Table 1, which shows a high positive correlation between asset volatility and capital across the sample within each period. (If the results for individual banks were assumed to be drawn from normal distributions for each variable, the 5 percent critical level for a null hypothesis of $\rho = 0$ would be 0.666; the correlation coefficients would be judged to be significantly greater than zero, except in 1985). The rise in market value capital ratios does not reflect simply a passive increase as bank stock prices increased along with the overall stock market. The book value of bank equity also rose, both absolutely and relative to the market value of bank assets, with much of the rise due to earnings retention and new equity issuance. Thus, the reduction in financial risk probably was an active response to a perceived need for greater capital.

Table 1
Correlation Between Risk and Capital

85:II	0.575	87:IV	0.948
85:IV	0.506	88:II	0.924
86:II	0.859	88:IV	0.724
86:IV	0.796	89:II	0.716
87:II	0.920	89:IV	0.714

Note: Coefficient of correlation between σ_A and k across banks, using case C1 assumptions.

VI. Conclusions

This paper considers the evolution of bank risk during the late 1980s, with a focus on nine large U.S. banking firms. The unusual element of this study is the use of exchange-traded options on bank holding company stocks to infer the volatility of bank assets and activities. The results show that operating risk increased by about 80 percent during the period. However, with the exception of the period around the 1987 stock market crash, financial risk generally declined. When the two separate changes are combined to examine overall risk to the deposit insurance fund, it appears that the burden imposed by large banks on the deposit insurer was little different at the end of the sample period than at the beginning.

Hence, while the data do support the notion that banks now engage in generally riskier activities than they did previously, market capital ratios have on average kept pace with the evolving mix of banking services and products, thus preventing deterioration in the degree of protection provided by bank capital. Large banks did not become substantially riskier in what is probably the most important public policy sense, the risk imposed on the deposit insurance fund. Additional regulatory actions to force reductions in bank risk-taking appear neither necessary nor warranted, at least on the strength of changes in risk at large banks during the late 1980s.

ENDNOTES

1. The only known previous use of these data (Levonian 1988) examined a shorter time period and used a simple but not strictly appropriate option pricing model.

2. The model presented by Levonian is a generalization of Merton's and two others, Marcus (1984) and Ronn and Verma (1986). The differences among the various models are summarized in Levonian (1991).

3. The main motivation for assuming that charter value is proportional to liabilities is modeling convenience; however, the assumption also is appropriate to the extent that franchise value is related to the size of a bank's "core" deposit base.

4. In a multiperiod setting, ϕB would reflect the discounted value of the future stream of rents as well.

5. This formulation differs slightly from Levonian (1991). In the present version, any regulatory costs are capitalized into the value of bank assets. Note that the dividend rate does not appear in (5) or (6). Equity is essentially a "dividend-protected" call option, in which the option holder receives the benefits of the dividend cash flow that otherwise would reduce the value of assets and hence equity.

6. The factor $N(x-\sigma_A\sqrt{T})$ is actually the probability that the bank would remain open in a world of risk-neutral investors, that is, a world in which the assets of the bank earn the risk-free rate of return.

7. According to Itô's Lemma, if $A(t)$ is determined by the stochastic process in equation (4), and equity is a function of $A(t)$ and t, then the differential of $E(A(t), t)$ is given by

$$dE = \frac{\partial E}{\partial A}\, dA + \frac{\partial E}{\partial t}\, dt + \frac{\partial^2 E}{\partial A^2}\, (dA)^2,$$

which is essentially a Taylor series expansion of E, with all higher-order terms vanishing as dt approaches zero (that is, in continuous time). Substituting for dA from (4), regrouping terms, and recognizing that $(dt)^2=0$, $dtdz=0$, and $(dz)^2=dt$ yields

$$dE = \left[\frac{\partial E}{\partial A}\,\mu_A A + \frac{\partial E}{\partial t} + \frac{\partial^2 E}{\partial A^2}\,\sigma_A^2 A^2\right] dt + \sigma_A A\,\frac{\partial E}{\partial A}\, dz.$$

The term in parentheses is the expected drift of the process. Defining μ_E as the expected drift term divided by E (to create a percentage rate of change), the differential dE can be written as

$$dE = \mu_E E dt + \sigma_E E dz,$$

where σ_E is given by (10). For a relatively simple discussion of Itô's Lemma and the associated stochastic calculus, see Haley and Schall (1979), Chapter 10. For a more rigorous but still accessible treatment, see Merton (1982).

8. This also is approximately true if the bank is very well-capitalized, so that the probability of closure is insignificant. As a bank moves closer to the point of closure, the contingent element becomes more important.

9. Again, if the contingent element of equity is trivial, then $N(x)\approx 1$ and $N'(x-\sigma_A\sqrt{T})\approx 0$, so that $\sigma_E\approx\sigma_A(A/E)$.

10. Linear regression analysis cannot rule out the possibility that c and vary in some nonlinear but known way during the sample period. Neither direct nor anecdotal evidence suggests that changes in either parameter are a major problem for the period studied, although King and O'Brien (1991) argue that regulators might systematically vary the threshold (and the monitoring interval) with the condition of each bank. Since this possibility cannot be ruled out, some degree of caution is appropriate in interpreting the results. Most previous studies using similar contingent-claim models have implicitly held both the closure threshold and the charter value ratio constant over time and across banks, often without the critical examination given to these assumptions in this paper; Furlong (1988) is an exception.

11. Schellhorn and Spellman (1991) calculate implicit volatilities from the prices of subordinated debt issued by bank holding companies, using the fact that risky debt can be valued as riskless debt minus a put option on the assets of the issuer. The spirit of the Schellhorn-Spellman approach is very similar to the present analysis. One drawback to their use of subordinated debt is that the sample size is smaller, because so few banks have regularly traded subordinated debt outstanding.

12. Day and Lewis (1988) make identical assumptions regarding interest rates and dividends.

13. Other methods for computing implied volatility typically involve weighted averages of observed option volatilities; see Bodurtha and Courtadon (1987, pages 28-30)

for a discussion. All of the methods give heavier weight to near-the-money options, and therefore in practice are likely to yield results close to those obtained through the simpler approach used here (arithmetically averaging only the nearest-the-money options).

14. The relatively small estimated insurance liability in Chart 5 is a result of the assumption that the FDIC sells the charters of failed banks to defray the cost of covering deposits; only the liability net of the charter value, as developed in equation (8), is presented.

REFERENCES

Barone-Adesi, G, and R. Whaley. 1987. "Efficient Analytic Approximation of American Option Values." *Journal of Finance* 42(3) pp. 301-320.

Black, F., and M. Scholes. 1972. "The Valuation of Options Contracts and a Test of Market Efficiency." *Journal of Finance* 27(2) pp. 399-417.

Board of Governors of the Federal Reserve System. 1985. "Revisions to Guidelines for Capital Adequacy." *Federal Reserve Bulletin* 71(6) pp. 440-441.

Bodurtha, J., Jr., and G. Courtadon. 1987. *The Pricing of Foreign Currency Options*. Salomon Brothers Center for the Study of Financial Institutions Monograph.

Day, T., and C. Lewis. 1988. "The Behavior of the Volatility Implicit in the Prices of Stock Index Options." *Journal of Financial Economics* 22(1) pp. 103-122.

Feinstein, S. 1988. "A Source of Implied Volatility Estimates." Working Paper 88-9. Federal Reserve Bank of Atlanta. (December).

Furlong, F. 1988. "Changes in Bank Risk-Taking." Federal Reserve Bank of San Francisco *Economic Review* (Spring) pp. 45-56.

Haley C., and L. Schall. 1979. *The Theory of Financial Decisions*. New York: McGraw-Hill.

Keeley, M. 1990. "Deposit Insurance, Risk, and Market Power in Banking." *American Economic Review* 80(5) pp. 1183-1200.

Kendall, S., and M. Levonian. 1991. "A Simple Approach to Better Deposit Insurance Pricing." *Journal of Banking and Finance* 15, pp. 999-1018.

King, K., and J. O'Brien. 1991. "Market-Based Deposit Insurance Premiums: An Evaluation." Federal Reserve Board of Governors (April) (unpublished).

Kwan, S. 1991. "Risk and Charter Value in Banking." Paper presented at the 27th Annual Conference on Bank Structure and Competition, May 1-3, Federal Reserve Bank of Chicago, Chicago, IL.

Latané H., and R. Rendleman. 1976. "Standard Deviations of Stock Price Ratios Implied in Options Prices." *Journal of Finance* 31(2) pp. 369-381.

Levonian, M. 1988. "The Value of F.D.I.C. Insurance Using Prices of Listed Options." Paper presented at the meeting of the Financial Management Association, October 12-15, New Orleans, LA.

________. 1991. "What Happens if Banks Are Closed 'Early'?" In *Rebuilding Banking: Proceedings of the 27th Annual Conference on Bank Structure and Competition*, May 6-8. Federal Reserve Bank of Chicago, Chicago, IL.

Marcus, A. 1984. "Deregulation and Bank Financial Policy. " *Journal of Banking and Finance* 8 (December) pp. 557-565.

________, and I. Shaked. 1984. "The Valuation of FDIC Deposit Insurance Using Option-pricing Estimates." *Journal of Money, Credit, and Banking* 16(4) pp. 446-460.

McDonald, R., and D. Siegel. 1984. "Option Pricing When the Underlying Asset Earns a Below-Equilibrium Rate of Return: A Note." *Journal of Finance* 39(1) (March) pp. 261-265.

Merton, R. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance* 29(3) pp. 449-470.

________. 1977. "An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees." *Journal of Banking and Finance* 1(3) pp. 3-11.

________. 1982. "On the Mathematics and Economics Assumptions of Continuous-Time Models." In *Financial Economics: Essays in Honor of Paul Cootner*, ed. W. Sharpe and C. Cootner, pp. 19-51. Englewood Cliffs: Prentice-Hall.

Neuberger, J. 1991. "Risk and Return in Banking: Evidence from Bank Stock Returns." Federal Reserve Bank of San Francisco *Economic Review* Fall pp. 18-30.

Ronn, E., and A. Verma. 1986. "Pricing Risk-Adjusted Deposit Insurance: An Option-Based Model." *Journal of Finance* 41(4) pp.871-896.

Schellhorn, C., and L. Spellman. 1991. "Subordinated Debt Market Information and the Pricing of Deposit Insurance." In *Rebuilding Banking: Proceedings of the 27th Annual Conference on Bank Structure and Competition*, May 6-8, Federal Reserve Bank of Chicago, Chicago, IL.

Schmalensee, R., and R. Trippi. 1978. "Common Stock Volatility Expectations Implied by Option Premia." *Journal of Finance* 32(1) pp.129-147.

Schwert, G.W. 1990. "Stock Volatility and the Crash of '87." *Review of Financial Studies* 3(1) pp.77-102.

Shaffer, S. 1991. "Aggregate Deposit Insurance Funding and Taxpayer Bailouts." *Journal of Banking and Finance* 15, pp. 1019-1040.

Smith, C. 1976. "Option Pricing: A Review." *Journal of Financial Economics* 3(1) pp. 3-51.

Risk and Return in Banking: Evidence from Bank Stock Returns

Jonathan A. Neuberger

Economist, Federal Reserve Bank of San Francisco. The author thanks Brantley Dettmer for research assistance. The members of the editorial committee, Brian Motley, Mark Levonian, and Gary Zimmerman, provided many helpful comments.

In this paper, I investigate the behavior of bank holding company stock returns from 1979 to 1990 in order to determine if bank risk has increased in recent years. Simple statistics on total volatility of returns indicate that the variance of bank stock returns rose in the latter part of the 1980s relative to earlier periods and to other stock and bond investments. In the context of equilibrium asset pricing models, I find that bank stock return covariance with respect to overall stock market returns increased during the 1980s while the sensitivity of bank stocks to returns on long-term debt securities declined. I also divide the sample by bank size and find that stocks of larger banks exhibited more stock market risk than smaller banks in the latter part of the 1980s, while no banks exhibited any statistically significant interest rate risk in the late 1980s.

There is currently a widespread perception that the U.S. banking system has become riskier in the past several years. The large number of bank failures, negative media coverage of the industry, and the rhetoric of legislative efforts in Washington to restructure the banking system all have contributed to this perception. Moreover, the legacy of the savings and loan crisis serves as a constant reminder of the excessive risks that some U.S. financial institutions undertook in the 1980s.

Industry observers have identified a number of factors that are potential causes of this apparent increase in bank risk. The usual list of suspects includes deregulation of financial markets, increased competition in banking, and financial innovation. The cause of any increased risk in banking probably will be a subject of debate for some time. Nevertheless, it is instructive to investigate the recent behavior of bank risk to determine if the public perception of greater risk is justified and whether any changes in risk have occurred in a systematic way. In this paper, I conduct such an investigation.

As a measure of bank risk, I consider the volatility of bank stock returns. Ideally, a direct measure of bank asset risk might be preferred, but it is difficult to observe the risks associated with specific bank assets. The behavior of bank stocks provides a reasonable, and readily available, alternative. In the absence of regulation or deposit insurance, there is a direct relationship between asset risk and stock risk. This relationship is complicated by the presence of financial regulation and the deposit insurance system, but the risk associated with holding bank stocks is still informative about the risks to the banking system. Moreover, the current focus on increased capital requirements for banks makes understanding bank stock risk particularly relevant. Common stock comprises the largest portion of bank capital and thus the value of bank equity provides a good proxy for bank net worth.

In this paper, I use a time series, cross-section sample of large U.S. commercial bank holding companies to examine the behavior of bank stock returns over the period 1979 to 1990. I consider first the overall volatility of these returns. Then, drawing on theories of capital asset pricing, I consider the influence of different systematic risk factors on the behavior of U.S. banks' stock returns.

The results from this analysis indicate that the relationship of bank stocks to systematic sources of risk in the economy has changed significantly during the past several years. Certainly, some sources of bank equity risk increased during the 1980s. However, my analysis shows that other sources of stock return variability actually declined during this period. My results also indicate that there is considerable variation among the banks in the sample regarding their equity risk. For example, I separate the banks in the sample by asset size. I find that the sensitivity of stock returns of large banks to overall stock market risk has increased relative to that of smaller banks. An understanding of such cross-sectional variations may help to identify potential winners and losers arising from proposals to reform the banking system.

While the study of bank stock returns provides useful insights into changes in bank risk, it is important to recognize the limitations of these data. The variability of bank stock returns reflects the market's perception of the risks associated with all aspects of bank holding company activities. These include asset risks, default risks, charter value risks, the risks associated with the value of the deposit insurance guarantee, and so on. It is not possible to infer from these data what has happened to a particular aspect of bank risk, say for example, to the riskiness of bank assets. (For a study of bank asset risk, see Mark Levonian's paper in this issue of the *Review*). The results here identify how the systematic risk factors included in asset pricing models influence the market's perception of this amalgam of bank stock risks and how this perception is reflected in bank equity returns.

I. Related Literature

According to the capital asset pricing model (CAPM, Sharpe 1964 and Lintner 1965), the return on a firm's equity can be explained as a function of a single factor, namely, the return on the market portfolio of assets. In empirical applications of the CAPM, the proxy for this market return typically is taken to be a broad measure of stock market returns (such as the S&P 500, or the AMEX composite index). The CAPM splits asset risk into two components. The first, called market or systematic risk, represents that portion of asset risk that is related to the riskiness of the market portfolio. The second component is called residual, or nonsystematic, risk and is the portion of total asset risk that is unrelated to the market portfolio. Because an investor can eliminate the effects of nonsystematic risk by suitably diversifying his portfolio, the CAPM argues that the expected returns on individual assets reflect only their systematic risk.

Bank stocks have been a frequent object of analysis in studies of equity risk and returns. Banks are of particular interest to economists because of their role as financial intermediaries. This role is believed to make bank stocks especially sensitive to changes in interest rates. To test this hypothesis, a number of studies have extended the basic CAPM formulation to include a measure of returns on debt in addition to the return on the market portfolio of stocks. This "two-index model" was first proposed by Stone (1974), and variations of this model have been investigated in subsequent work by Martin and Keown (1977), Lloyd and Schick (1977), Lynge and Zumwalt (1980), Chance and Lane (1980), Flannery and James (1982, 1984a, 1984b), Beebe (1983), and Booth and Officer (1985). With the exception of the paper by Chance and Lane, these studies have found that bank stock returns exhibit sensitivity to interest rates over and above their sensitivity to stock market changes. Moreover, this sensitivity exceeds that shown by most nonfinancial firms, confirming the notion that the particular nature of bank assets and liabilities makes them especially sensitive to changes in interest rates.[1]

A number of studies have attempted to relate the market and interest rate sensitivities of bank stock returns to some aspect of bank balance sheet composition. Dietrich (1986), for example, argues that the estimated coefficients in the two-index model should depend on the balance sheet proportions of broad categories of bank assets and liabilities. He embeds this hypothesis into the two-index model and estimates portfolio composition effects on the risk factors. Dietrich finds that market risk, the estimated coefficient on the market portfolio of stocks, is most heavily influenced by lending activity, time deposits, and long-term debt relative to assets. Interest rate risk, he finds, is most affected by the proportion of time deposits in the balance sheet. While Dietrich's results suggest that balance sheet composition may be important in explaining the risk characteristics of bank stocks, his empirical results suffer from serious econometric problems. The asset and liability categories used in that study also are too broad to be of much practical use in identifying specific sources of bank risk.

In a similar avenue of research, Rosenberg and Perry (1981) consider the determinants of bank risk in a single-index CAPM. More specifically, they estimate the effects on systematic and residual risk of a large number of asset and liability ratios, operating characteristics, stock market variables, and regional indicators. These authors find that a number of their chosen indicators are significant determinants of bank risk. More interesting, they find that different indicators help to explain systematic and nonsystematic risk of bank stock returns. For example, size, dividend yield, equity capital, and the asset/long-term liability ratio all help to predict market-related risk. For residual risk, earnings variability and leverage are the most important determinants. Rosenberg and Perry suggest that bank risk can be predicted by focusing on a few significant indicators, and that efforts to understand bank risk should focus on understanding these aspects of bank behavior.

One weakness of the studies cited above is that they provide little theoretical justification for the particular empirical specifications used. A study by Flannery and James (1984a) relies on a firmer theoretical foundation for the analysis of bank risk and return. In this work, the authors test the so-called "nominal contracting hypothesis" (French, Ruback, and Schwert 1983) on a sample of bank and thrift stocks. This hypothesis suggests that a firm's holdings of nominal assets and liabilities affect its common stock returns through the redistributive effects of unanticipated inflation and unanticipated changes in expected inflation. More specifically, the nominal contracting hypothesis suggests that the interest rate sensitivity of a firm's stock will be larger the greater the amount of net nominal assets (that is, nominal assets minus nominal liabilities) and the longer the duration of those net nominal assets.

Flannery and James first estimate a two-index model of stock returns on a time series, cross-section sample of bank stocks. They then develop a proxy for the duration of a bank's net nominal assets and regress the estimated interest rate coefficients on this duration measure. Nominal asset duration is highly significant in explaining the size of the interest rate sensitivity of bank stock returns. Kwan (1991) extends this work by estimating the Flannery and James model simultaneously in a random coefficients framework. These studies confirm that the composition of a bank's balance sheet, here as measured by the duration of its net nominal assets, can influence the sensitivity of bank stock returns to changes in interest rates.

II. Current Modeling Approach

Two Models of Asset Pricing

As the discussion of the related literature shows, most researchers have employed a particular empirical model of capital asset prices in order to focus on some aspect of bank stock risk. Some debate persists among economists as to the "correct" specification to use for describing equity returns. In this paper, I investigate the behavior of bank stocks in the context of two different models of asset pricing: the single-index CAPM and a two-factor model.

The typical CAPM formulation is specified as follows:

$$(1) \qquad R_{jt} = \alpha_j + \beta_{Mj} R_{Mt} + \epsilon_{jt},$$

where R_{jt} is the expected excess holding period return on the equity of bank j in period t, R_{Mt} is the expected excess holding period return on the market portfolio of stocks in period t, β_{Mj} is a parameter to be estimated that represents the sensitivity of the stock of bank j to changes in the expected return on the market portfolio, α_j is another estimated parameter that indicates deviations from equilibrium pricing, and ϵ_{jt} is the residual left unexplained by the expected return on the market portfolio.[2]

The parameter β_{Mj} is a measure of the covariance of return on an individual stock with the return on the market portfolio of risky assets. It thus represents the sensitivity of that stock to systematic, or nondiversifiable, risk.[3] According to the CAPM, an "average" stock in the market portfolio will have a value of β_{Mj} equal to one. An asset with β_{Mj} greater than one carries above average nondiversifiable risk, and must provide a greater than average expected return in order to induce investors to hold it. The CAPM predicts that only nondiversifiable risk is rewarded by a higher expected return. Risk that is idiosyncratic to the individual stock, and can therefore be diversified away, is not associated in equilibrium with higher expected returns. The model thus predicts that the expected value of α_j is zero. Of course, realized or *ex post* values of α_j can differ from zero if new information affects the asset's price and return during the estimation period.

As mentioned in the previous section, the primary hypothesis underlying the CAPM is that the return on the market portfolio is a sufficient statistic to determine the return on individual assets. One implication of this model, therefore, is that no other variables should be significant in explaining asset returns. Empirically, this prediction often has not been supported, leading to asset pricing models in which additional or alternative factors have been included to capture missing influences on individual asset returns.

The two-factor model augments the CAPM by adding as an additional explanatory variable the expected return on a debt security. The logic behind this model, as first proposed by Stone (1974), is that investors have two general categories of assets to choose from: equity shares and debt securities. As a result, expected returns on both types of instruments should be relevant in setting the price of individual financial assets. This same type of two-factor model also can be derived more rigorously from Merton's (1973) intertemporal version of the CAPM, or from Ross's (1976) multi-factor arbitrage pricing theory. A number of tests of the two-factor model using stock returns of industrial companies found little significance for debt returns. However, stocks of companies in certain sectors, such as utilities and financial intermediaries, typically exhibit significant sensitivity to changes in returns on debt securities.

The two-factor model takes the form

$$R_{jt} = \alpha_j + \beta_{Mj} R_{Mt} + \beta_{Ij} R_{It} + \epsilon_{jt}, \tag{2}$$

where R_{It} is the expected excess holding period return on a selected debt security in period t, β_{Ij} is a parameter that captures the sensitivity of asset j to changes in the expected holding period return on debt, and the other variables and parameters are as defined in equation (1).[4]

Two modeling issues arise in empirical applications of the asset pricing equations (1) and (2). First, time-series regressions of these equations imply that the estimated coefficients should be constant over time. Evidence suggests, however, that estimated βs exhibit considerable temporal variation. Moreover, efforts to relate the estimated coefficients to balance sheet composition variables suggest that these coefficients will change with changes in the asset/liability mix of banks. Recent evidence by Kane and Unal (1988) using a switching regression methodology and by Kwan (1991) in the context of a random coefficients framework confirms the nonstationarity of the debt return β in equation (2). Other work in a CAPM framework likewise suggests that β_{Mj} varies over time. In order to deal with this issue, I estimate versions of the two asset pricing equations over different subsamples of the 1979 to 1990 period. This procedure generates statistics that enable me to test for the constancy of the estimated regression coefficients.[5]

The second modeling issue involves possible multicollinearity between the two returns series used as explanatory variables in equation (2). Chance and Lane (1980) argue that returns on debt probably are influenced by the same factors that determine the returns on the market portfolio of stocks. One way to deal with this issue, they suggest, is to orthogonalize one of the series by regressing it on the other. The residual series from this regression, which by definition is uncorrelated with the other explanatory variable, then can be used as a regressor in the equity returns estimating equation. This procedure eliminates the estimation bias and isolates stock market from extramarket effects on stock returns. Of course, the direction of causality in this first-stage regression is not clear. Chance and Lane regress the debt return variable on the stock market return while others, including Lynge and Zumwalt (1980) and Flannery and James (1982), perform the opposite regression.

This second issue may be important for hypothesis testing. Giliberto (1985) shows that the estimated standard errors of the second-stage regression coefficients are unbiased only for the series that was used as the dependent variable in the first-stage regression. This means that studies regressing the stock market index on the debt returns variable, like Flannery and James (1982), may produce biased estimates of the significance of β_{Ij}, but will yield unbiased estimates of the standard errors of β_{Mj}. To determine the empirical significance of this bias, I reestimated all of the equations presented in the next section using both orthogonalizations. While the two series did exhibit significant cross correlation, the orthogonalized results did not differ in a statistically significant way from those reported here.[6] This suggests that any bias resulting from the multicollinearity between the explanatory variables in the two-factor model is not substantial enough to alter the empirical results.

The two asset pricing models described above predict that different firms' equity returns will exhibit differing sensitivities to stock market and debt returns. In terms of the models' parameters, this means that each firm will have its own specific values of α_j, β_{Mj} and β_{Ij}. The estimation results described in the next section are from pooled regressions that combine time-series observations from all the banks in the sample. In Section IV, I group the banks into four size categories and describe regressions on these subsets of banks. The estimated parameter values presented in Tables 2 and 3 are thus averages of the αs and βs for different samples of banks. In this paper, I do not present estimated parameters for individual banks. To reflect this "averaging" in the discussion below, I drop the j subscripts from all α and β parameters (referring to individual banks) when describing the estimation results.

Data

In the current study, I estimate monthly stock returns of a sample of 84 large bank holding companies taken from the Compustat Bank tapes. The monthly returns are derived from end-of-month stock prices, and are adjusted for dividends and stock splits. The Compustat tapes include data on bank holding companies whose stocks trade on a major exchange. This means that the sample includes primarily large banks and is thus not completely representative of all U.S. banks. Of the 84 bank holding companies considered in the current study, the smallest held assets in the first quarter of 1990 of $2.3 billion, while the largest had over $230 billion in assets. The sample was chosen on the basis of availability over the entire interval 1979 to 1990. This period provides a number of observations prior to deregulation of bank interest rates, and also encompasses several cyclical episodes. With 144 time points and 84 banks, the sample includes over 12,000 data observations.

The return on the market portfolio is proxied by the monthly return series on the equal-weighted Standard & Poor's 500 index of stocks. This variable was taken from the Center for Research in Securities Prices (CRSP) tape for the period 1979 to 1988, and from DRI's U.S. Central database for 1989 and 1990. The debt return series is an approximation of the monthly holding period return on 30-year constant maturity U.S. Government bonds. The approximation, as suggested by Flannery and James (1984b), is

$$(3) \qquad -\frac{(Y_t - Y_{t-1})}{Y_{t-1}},$$

where Y_t is the investment yield in month t on the bond. This expression is the percentage change in the bond's yield, multiplied by -1. Note that monthly bond returns fall as yields rise. Thus, a positive estimated coefficient on this variable implies that bank stock returns are negatively affected by increases in bond yields. The yield series used in the construction of this variable was obtained from Citibase.

III. Bank Stock Risk and Return Over Time

A First Look

In describing changes in bank stock risk, it is essential to have an accepted measure of that risk. From the standpoint of portfolio theory, expected or *ex ante* risk is the relevant factor that determines asset prices. Unfortunately, such anticipated risk is generally unobservable. As a proxy, it is common to look at realized, or *ex post*, risk as the appropriate descriptor of asset risk, with the belief that past volatility is a likely indicator of future volatility.[7] Economists typically consider the total variance of historical asset returns (or its square root, the standard deviation) as an appropriate measure of the overall volatility associated with asset returns.

Table 1 contains summary statistics of monthly holding period returns for three different groups of assets for the 1979-1990 period, as well as for four subperiods of that interval. In the first two columns, I present the period averages of the mean and standard deviation of monthly returns for the sample of 84 bank holding company stocks. The second pair of columns contains comparable statistics for the 30 Dow Jones Industrial firms. In both cases, the numbers presented in Table 1 are unweighted averages of the

Table 1

Summary Statistics for Monthly Asset Returns, 1979 to 1990

Interval	84 Bank Stocks (unweighted average)		30 Stocks in Dow Jones Industrials (unweighted average)		30-Year U.S. Treasury Bonds	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Whole-Sample Period						
1979:2-1990:12	0.012	0.083	0.015	0.078	0.001	0.037
Quarter-Sample Subperiods						
1979:2-1981:12	0.015	0.073	0.007	0.070	−0.010	0.045
1982:1-1984:12	0.020	0.078	0.017	0.078	0.005	0.033
1985:1-1987:12	0.012	0.087	0.017	0.091	0.007	0.040
1988:1-1990:12	0.001	0.087	0.019	0.064	0.003	0.025

individual company means and standard deviations during the applicable period.[8] The last two columns in Table 1 contain the mean and standard deviation of the monthly return on 30-year constant maturity Treasury bonds.

The first row of Table 1 contains statistics for the 12-year period, 1979 to 1990. Over this interval, the mean monthly return on both groups of stocks significantly exceeded the mean return on bonds. At the same time, the total risk associated with holding either of these groups of stocks was more than twice the risk of holding Treasury bonds. Between the two samples of stocks, the 30 industrial firms provided a slightly higher mean monthly return and faced somewhat less average risk than the sample of bank holding company stocks, although the differences between the two groups are small. During the full 12-year period, it does not appear that bank stocks were significantly riskier than other equities.

In the bottom portion of Table 1, I divide the full sample period into four subperiods and present the averages of mean monthly returns and standard deviations for the three groups of assets during these different subperiods. The 30 industrial stocks show an upward trend in returns over the four subperiods of the sample, while the bank stocks exhibit a generally downward trend. Notably, in only the 1988-90 subperiod were bank stock returns below the returns on both the 30 industrial stocks and the T-bonds.

While bank stock returns declined in the latter half of the sample period, the variance of these stock returns rose over the course of the 12-year sample period. The average standard deviation of return on the 84 bank holding company stocks was 20 percent higher in the last two subperiods than it was in the first part of the sample. The average risk of the 30 industrial stocks rose through the 1985-87 period, but then declined in the 1988-90 period. The standard deviation of bond returns fluctuated during the four quarters of the sample without any apparent trend. Again, it is notable that, in the last subperiod of the sample, the standard deviation of bank stock returns exceeded the standard deviations of the other two groups of assets. Thus, by the end of the 1979-90 period, bank holding company stocks were more volatile than the other assets and offered investors a lower rate of return.

The numbers presented in Table 1 provide support for the perception that bank stocks have become riskier in recent years. The volatility of bank holding company stock returns increased during the 1980s, both in absolute terms and relative to other portfolios of financial assets, including other equities. At the same time, the average returns to holding bank stocks declined significantly. By the latter half of the 1980s, it appears that investors in bank equities suffered from both higher risk and lower returns.

Risk in the Context of the Two Asset-Pricing Models

The summary statistics of Table 1 confirm that the total variability of bank stock returns increased over the 1979 to 1990 period. However, it is useful to determine if the sensitivity of bank stocks to systematic sources of risk changed during this period. Finance theory predicts that (expected) asset returns should depend on systematic risk and not on total risk. For example, the CAPM suggests that only risk associated with returns on the market portfolio will be compensated by higher expected asset returns. Similarly, the two-factor model predicts that market risk and interest rate risk should be compensated by higher returns. Thus, the higher risks and lower returns on bank stocks observed in the bottom portion of Table 1 could still be consistent with equilibrium asset pricing models if returns fell because systematic risk declined. Estimation of equations (1) and (2) in the previous section can help to shed light on this point.

Table 2 contains regression results from estimating the CAPM and the two-factor model on time series of the monthly stock returns of the 84 bank holding companies in the sample. The coefficients from these estimates correspond to equations (1) and (2) discussed above. The parameter estimates presented in Table 2 are average values across the 84 bank holding companies in the sample. The top part of Table 2 contains estimation results for the full 12-year sample period, while the bottom portion of the table contains estimates from the four subperiods. I test the significance of the estimated coefficients against the null hypotheses that α and β_I are zero and β_M is one.

The CAPM results for the whole sample show that the average covariance of bank holding company stocks relative to the S&P 500 index was less than the "average" stock during the 1979-90 period, as indicated by the estimated β_M of 0.92, which is significantly less than one. This suggests that, over the 12 years of the sample interval, changes in the stock market as a whole were associated with less than one-for-one changes in bank stocks. A long-run value of β_M that is close to one is reasonable because banks are expected to hold diversified portfolios of loans and other assets whose returns should mimic the behavior of the broader market. While this may not be true for small, regional banks, it certainly should apply to the relatively large banks included in the current sample. The positive, significant value of α suggests that, on average, the sample of bank holding company stocks was underpriced during the 12-year period, yielding returns in excess of what would be predicted by the basic CAPM. The model explains only about 25 percent of the total variance of returns during the sample period. This means that bank stock returns contained a large portion of nonsystematic risk.

In the two-factor model for the 12-year sample period, both factors were highly significant in explaining bank stock returns from 1979 to 1990. This confirms previous evidence regarding the sensitivity of bank equity returns to changes in bond yields over and above their sensitivity to the stock market. Moreover, the estimated coefficient, β_I, is positive, indicating that bank stock returns were negatively affected by increases in long-term yields during the sample period. Adding the debt returns variable to the estimating equation reduces somewhat the stock market sensitivity of bank equities. While the change in this coefficient suggests that the two factors may be collinear, the results were the same when the explanatory variables were purged of their common influence. As in the CAPM formulation, the positive and significant value of α suggests that bank stocks were, on average, underpriced during the 12-year period. The expanded model explains about 27 percent of the variance of bank stock returns, only slightly better than the CAPM, and again indicating substantial nonsystematic risk.

Several striking results stand out from the estimates of the two models for the subperiods. First, in the context of

Table 2

Results from One- and Two-Factor Models of Monthly Bank Stock Returns

Model	Interval	α	β_M	β_I	RMSE	$\bar{R}^2$
Whole-Sample Regressions						
CAPM	1979:2-90:12	0.003*** (0.0007)	0.916*** (0.0146)		0.073	0.247
Two-factor	1979:2-90:12	0.005*** (0.0007)	0.828*** (0.0150)	0.373*** (0.0182)	0.071	0.272
Quarter-Sample Regressions						
CAPM	1979:2-81:12	0.007*** (0.0012)	0.719*** (0.0280)		0.067	0.183
Two-factor	1979:2-81:12	0.018*** (0.0013)	0.604*** (0.0266)	0.547*** (0.0259)	0.062	0.291
CAPM	1982:1-84:12	0.011*** (0.0013)	0.929*** (0.0305)		0.070	0.235
Two-factor	1982:1-84:12	0.013*** (0.0013)	0.716*** (0.0354)	0.510*** (0.0452)	0.069	0.266
CAPM	1985:1-87:12	−0.001 (0.0012)	0.928*** (0.0207)		0.068	0.400
Two-factor	1985:1-87:12	−0.001 (0.0012)	0.916*** (0.0205)	0.259*** (0.0311)	0.067	0.413
CAPM	1988:1-90:12	−0.008*** (0.0015)	1.297*** (0.0488)		0.082	0.189
Two-factor	1988:1-90:12	−0.008*** (0.0015)	1.256*** (0.0657)	0.075 (0.0793)	0.082	0.189

Note: *** = significant at 99 percent confidence level
** = significant at 95 percent confidence level
* = significant at 90 percent confidence level
Significance tests are against null hypotheses that the estimated coefficients are zero for α and β_I, and one for β_M.
Standard errors are in parentheses.

the CAPM, market-related systematic risk of bank equity returns, as embodied in β_M, increased during the four subperiods of the 12-year interval. The estimated value of β_M rose from 0.72 in the 1979-81 period to 1.30 during 1988-90.[9] Investors who held bank equities faced more market-related risk over the period and were rewarded for assuming this additional systematic risk by receiving a higher return.

The increase in market-related risk is even more striking when viewed in the context of the two-factor model. Estimated values of β_M more than doubled from the beginning to the end of the sample period, from 0.60 in 1979-81 to 1.26 in 1988-90.[10] These estimates confirm that the systematic, market-related risk of bank holding company stocks increased dramatically during the 1979 to 1990 period.

Perhaps the most striking result in the quarter-sample estimates is the progression of the estimated coefficients on debt returns. The values of β_I decrease monotonically over the four subperiods of the sample, from above 0.50 during both the 1979-81 and 1982-84 periods, to 0.26 from 1985 to 1987, to insignificantly different from zero during the 1988-90 sample period. In contrast to stock market-related risk, the sensitivity of bank equity returns to bond yields declined during the past decade. Moreover, bank holding company stock returns showed no sensitivity to changes in yields in the last three-year period of the sample, the only subinterval for which this was true. While bank stocks faced greater volatility with respect to the stock market, they clearly became increasingly insulated from the effects of bond yield changes.

Of course, systematic bank stock risk is only one aspect of total risk. The remaining portion of risk represents residual, or nonsystematic, bank equity risk. This, too, changed significantly during the 1980s. The two-factor model, for example, explains between 25 and 40 percent of total bank stock returns during the first three subperiods of the 12-year sample interval. By the 1988-90 period, this model explains less than 20 percent of total returns. Thus, the model leaves a large component of bank equity returns unexplained. Clearly, bank stocks entail a substantial amount of asset-specific risk that is not accounted for by the systematic risk factors of these asset pricing models.

Finally, while bank stocks apparently were underpriced on average during the 12-year period, as indicated by the positive values of α in the first two rows of Table 2, market pricing of bank stocks changed during the course of the 1980s. Estimated intercepts were significantly positive during the first two quarters of the sample period, were statistically indistinguishable from zero in the 1985-87 period, and then turned significantly negative in the last part of the sample interval. In terms of the asset pricing models estimated in Table 2, this means that the stocks of the 84 bank holding companies were overpriced in the 1988-90 period, yielding a lower return than the models would predict.

IV. Some Cross-Sectional Comparisons

While the estimates presented in Table 2 contain important information about bank equity risks during the past 12 years, they also conceal substantial cross-sectional variation in bank stocks' responses to stock market and interest rate risk. For example, I estimated values of β_{Mj} and β_{Ij} for each of the 84 bank holding companies during the various subintervals of the sample period. I then generated summary statistics on these "samples" of coefficient estimates. The variance of these coefficients was by far the greatest in the 1988-90 period. That is, there was considerably more variation across bank stocks in their stock market and bond yield sensitivity during the last three years of the 12-year sample period than in any other part of the interval. This suggests that banks may have responded in different ways to changes in their economic and regulatory environment.

One way to separate banks in the sample is by size. It is reasonable to assume that the stocks of different-sized banks may face differing sensitivities to systematic risk factors. For example, the stock of a large bank holding company may reflect its enhanced opportunities for asset diversification. Such a large bank thus may exhibit less variability relative to the broader market than a smaller bank whose opportunities for diversification may be more limited. Similarly, a large bank may be able to exploit possible economies of scale in hedging against interest rate risk that a small bank cannot. These differences will show up in the asset pricing models in terms of different values of β_I. Moreover, if regulators implement, either explicitly or implicitly, a policy of protecting large banks from failure while permitting smaller institutions to go under, such a policy may be reflected in estimates of the asset pricing models and probably will differ across institutions.

To address this question, I split the sample of 84 bank holding companies into four size categories according to dollar amount of assets as of the first quarter of 1987. I then estimated separate regressions for each category.[11] The estimates follow a distinct pattern where size clearly is relevant to the banks' stock sensitivities to the two risk factors. To highlight the differences between the various groups, I present in the four panels of Table 3 the subsample results from the two-factor model for four sizes of banks: assets less than \$5 billion (13 banks), assets between \$5 and \$20 billion (37 banks), assets between \$20 and \$55 billion (24 banks), and assets greater than \$55 billion (10 banks). As in Table 2, significance levels test against a value of β_M equal to one and values of α and β_I equal to zero.

The results in Table 3 suggest that the greatest differences in estimated parameters are between the stocks of the smallest banks in the sample and those of the remaining banks; the three larger categories of banks show quite similar estimation results. For example, the three groups of larger banks all exhibited generally increasing values of β_M over the course of the four subperiods. In contrast, there is no clear trend to the estimated values of β_M for the smallest banks in the sample. Thus, the stocks of the larger banks all became more sensitive to stock market-related risk during the 1980s, while the smaller bank stocks showed no tendency to entail higher market risk. It is notable that the smaller bank stocks had the highest market risk in the first portion of the sample period, 0.8 versus values of β_M between 0.4 and 0.6 for the other three categories of banks. By the end of the sample period, however, the smallest banks had by far the lowest estimated values of β_M, 0.7. The other groups of banks all had estimates of β_M in the last period that exceeded one (although only the largest two groups had parameter estimates that were significantly different from one). Moreover, the stocks of the largest group of banks exhibited the greatest sensitivity to stock market risk of any banks in the sample. The estimated parameter value of 1.8 in the last period of the sample is larger than any other point estimate in this study.

Bank stock sensitivity to bond yields also showed an interesting pattern. Again, the stocks of the smallest group of bank holding companies stand out from the others, while the other three groups of bank stocks look very similar. The stock returns of the three groups of larger banks all exhibited significant sensitivity to bond yields in the first subperiod of the sample, with estimated values of β_I near 0.6. In contrast, the group of smaller bank stocks showed little sensitivity to yield changes, as indicated by the coefficient estimate of 0.2. As the 1980s progressed, the sensitivity of the stock returns for the three categories of larger banks all declined until, in the final three-year period of the sample, none of the banks' equity returns showed any evidence of significant interest rate risk. The stocks of the smallest banks in the sample continued to exhibit little or no interest rate risk in the four subperiods of the sample. While the point estimates remain about the same (0.2), the estimated standard errors increase over time such that the coefficient on the debt return variable is statistically insignificant in the last portion of the sample.

The $\bar{R}^2$ statistics from these regressions indicate that the estimates for each size group leave a large portion of bank stock returns unexplained. Thus, stocks of the different-sized banks in the sample all have a significant component of nonsystematic risk. Moreover, the $\bar{R}^2$ for all four groups declines in the last part of the sample interval, indicating that the proportion of bank stock returns attributable to the two systematic risk factors fell in the 1988-90 period. This is particularly true for the smallest banks in the sample. While the two-factor model explained about 20 percent of stock returns for the other three size groups from 1988 to 1990, it provided less than 10 percent of the explanation for the smallest group of banks. It is not surprising that the stocks of the smaller banks in the sample exhibited the most nonsystematic risk since these smaller banks may be more heavily influenced by bank-specific events and local market conditions. Nevertheless, all banks in the sample, including the largest ones, exhibited significant nonsystematic equity risk.

Finally, the estimated values of α follow the same pattern as for the entire sample, and are roughly similar for all size categories of banks. αs are positive in all four cases early in the 12-year sample period, and all turn negative in the last subperiod. As mentioned above, this means that bank stocks provided abnormally high returns (relative to the predictions of the theoretical asset pricing models) in the late 1970s and early 1980s, and abnormally low returns in the late 1980s.

Table 3

Results From Two-Factor Model by Bank Size
Quarter-Sample Regressions

Interval	α	β_M	β_I	RMSE	$\bar{R}^2$
Panel 1: Banks with Assets less than $5 billion—13 Banks					
1979:2-81:12	0.016***	0.809**	0.186**	0.069	0.233
	(0.0035)	(0.0750)	(0.0729)		
1982:1-84:12	0.002	0.601***	0.239*	0.073	0.143
	(0.0034)	(0.0953)	(0.1218)		
1985:1-87:12	−0.009**	0.844**	−0.004	0.083	0.270
	(0.0039)	(0.0641)	(0.0969)		
1988:1-90:12	−0.008*	0.729	0.266	0.084	0.091
	(0.0040)	(0.1722)	(0.2076)		
Panel 2: Banks with Assets between $5 billion and $20 billion—37 Banks					
1979:2-81:12	0.017***	0.589***	0.604***	0.060	0.323
	(0.0018)	(0.0385)	(0.0374)		
1982:1-84:12	0.018***	0.571***	0.626***	0.064	0.271
	(0.0017)	(0.0494)	(0.0631)		
1985:1-87:12	0.003	0.886***	0.254***	0.065	0.415
	(0.0018)	(0.0298)	(0.0835)		
1988:1-90:12	−0.006***	1.143	0.150	0.073	0.206
	(0.0021)	(0.0887)	(0.1070)		
Panel 3: Banks with Assets between $20 billion and $55 billion—24 Banks					
1979:2-81:12	0.020***	0.606***	0.627***	0.060	0.336
	(0.0023)	(0.0480)	(0.0466)		
1982:1-84:12	0.013***	0.859**	0.452***	0.072	0.288
	(0.0025)	(0.0689)	(0.0880)		
1985:1-87:12	0.002	0.922**	0.347***	0.062	0.464
	(0.0021)	(0.0354)	(0.0536)		
1988:1-90:12	−0.014***	1.490***	0.012	0.086	0.218
	(0.0030)	(0.1294)	(0.1560)		
Panel 4: Banks with Assets greater than $55 billion—10 Banks					
1979:2-81:12	0.019***	0.394***	0.616***	0.064	0.236
	(0.0037)	(0.0791)	(0.0769)		
1982:1-84:12	0.010***	1.057	0.571***	0.069	0.399
	(0.0037)	(0.1031)	(0.1318)		
1985:1-87:12	−0.010***	1.104*	0.407***	0.062	0.550
	(0.0033)	(0.0552)	(0.0835)		
1988:1-90:12	−0.001	1.798***	−0.304	0.094	0.217
	(0.0051)	(0.2189)	(0.2639)		

See Notes to Table 2.

V. Conclusion

The results presented above highlight a number of interesting aspects of the behavior of bank holding company stock returns from 1979 to 1990. For example, the total variability of bank equity returns increased during the sample period relative to the returns on industrial equities and on bonds. Moreover, this increased total volatility of returns occurred at the same time that the level of average bank equity returns fell relative to the other assets. By the end of the 1980s, holders of bank stocks faced relatively higher risk and relatively lower returns.

In the context of the asset pricing models estimated in this paper, changes in the total risk and return of bank equities were accompanied by a significant shift in the sensitivity of bank stocks to systematic risk factors. The covariance between bank equity returns and a broad stock market index definitely rose on average during the 1980s. In the latter part of the 1980s, average values of stock market betas for the 84 bank holding companies in the sample exceeded one. Thus, changes in returns on the S&P 500 stock index were associated with a greater than one-for-one movement in bank stock returns, whereas they were less than one-for-one in the early 1980s. This increased stock market sensitivity was especially pronounced for the larger banks in the sample. Thus, the stock returns of large bank holding companies became increasingly sensitive to factors that influence the overall stock market.

One of the most striking findings in this paper is the decline in the bond yield sensitivity of bank stock returns during the estimation period. By the last three-year period in the sample, banks stocks showed no statistically significant evidence of any effects of bond yields on their returns. Moreover, this finding was consistent across banks of all sizes in the sample. The recent lack of bond yield sensitivity contrasts sharply with the behavior of the same bank stocks in the earlier part of the sample period as well as with the findings of previous studies. It is possible to interpret this reduction in interest rate risk as the result of bank managers making greater use of adjustable rate instruments and other hedging strategies to insulate their stock returns from the effects of changes in yields. It is reasonable to conclude that interest rate deregulation and financial market innovations, such as interest rate swaps, financial futures contracts, and adjustable rate mortgages, helped to reduce the interest rate risk of bank stocks by widening the sphere for banks to engage in risk hedging activities.

Of course, there may be alternative explanations for the apparent lack of interest rate risk in bank stock returns in the last part of the 1980s. Shifts in the observed sensitivity of bank equities can reflect changes not only in bank behavior but also in the regulatory environment in which banks operate. For example, in the late 1980s, bank regulators from around the world were negotiating the structure of international risk-based capital standards under the aegis of the Bank of International Settlements. By 1987, the likely future shape of these standards was becoming clear. Under the new standards, risk adjustments to regulated capital levels would be made on the basis of credit risk only and would downplay interest rate risk. While banks might be expected to respond to this change in regulation by increasing their interest rate risk exposure, the change in the enforcement policies of regulators could attenuate the impact of such actions on bank equity values. The net result could be a reduction in the interest rate risk embedded in bank stock returns.

Alternatively, the observed reduction in the debt return sensitivity of bank stocks might be partially explained by a statistical phenomenon. If the variance of debt returns fell significantly from 1988 to 1990 while bank stock returns behaved similarly to earlier subperiods, this might explain the lack of significance for the coefficient on debt returns in the last part of the sample. In fact, the variance of the debt returns series did fall somewhat in the last three years of the sample relative to earlier subperiods. However, this drop in variance probably was not large enough by itself to account for the dramatic decline in the estimated β_I values from 1988 to 1990. More likely it is a combination of factors related to changes in bank behavior, regulatory shifts, and statistical effects that contributed to reduce the measured sensitivity of bank stock returns to changes in bond yields in the last part of the sample period.

Finally, the results presented above support the conclusion that the proportion of nonsystematic risk in bank stocks rose during the 12-year sample period. The asset pricing models explain at most 40 to 50 percent of stock return variability for certain size categories of banks during certain subsamples of the estimation period. The proportion of total variance explained by the systematic risk factors declined for the total and for each size group of banks in the last three-year period of the sample. This means that, in the late 1980s, bank stock risk was more related to bank-specific factors than at any other time since the late 1970s. The increase in nonsystematic risk was greatest for the smaller banks in the sample. An accurate assessment of the stock risk of these banks thus requires less consideration of systematic risk factors and more careful attention to factors specific to the individual institutions.

ENDNOTES

1. The two-index model, though proposed by Stone (1974) in a somewhat *ad hoc* fashion, also can be derived formally from the intertemporal CAPM of Merton (1973), as well as from the Arbitrage Pricing Theory (APT) of Ross (1976). The latter framework suggests that there may be additional factors besides the two considered in these studies that are relevant to explaining asset returns. For example, Chen, Roll and Ross (1986) derive an APT model in which five prespecified macroeconomic factors are used to explain returns on several portfolios of stock. These authors find that several of the factors are important in explaining the returns on diversified stock portfolios. Campbell, Dietrich and Weinstein (1985) test the significance of these five factors on portfolios of financial intermediary stocks. They find that banks stocks are particularly sensitive to measures of default risk and term structure premia (both related to interest rates) as well as to the stock market index. These findings provide some support for the two-index formulation used so extensively in the banking literature.

2. The two holding period return series are expressed in the CAPM in terms of their return in excess of the return on a risk-free security, usually assumed to be a short-term riskless government bond. If no asset is considered risk-free, then it may be possible to construct an asset whose rate of return has zero covariance with the market portfolio. In this "zero-beta" version of the CAPM, the return on this security is considered to be the risk-free rate of return. See Fama (1976) for discussion of this point.

3. If changes in the stock market, and thus the return on the market portfolio, mirror movements in the macroeconomy, then the stock market beta can also be interpreted as measuring the sensitivity of the asset's return to changes in general economic conditions.

4. Again, all holding period returns are expressed in excess of the risk-free rate of return, where that rate is the yield on a short-term Treasury bill.

5. There are, of course, alternative ways to test for time-varying effects on the estimated coefficients. For example, a time trend could be included as an explanatory variable in the regressions, although this would constrain time effects to be linear and monotonic over the estimation interval. Alternatively, it is possible to estimate a shift parameter by interacting dummy variables for different time periods with the explanatory variables in the regression. Beebe (1985) uses this methodology. Another method is to assume that the estimated coefficients depend on some time-varying factor. Embedding this assumption into the regression equations translates into including additional, interacted explanatory variables in the estimates. See Kwan (1991) for an example of this latter approach. The methodology adopted here provides considerable flexibility without imposing additional theoretical or empirical constraints, and generates easily interpreted test statistics.

6. The correlation coefficient between stock market and debt returns was 0.28 during the 12-year sample period and was significantly different from zero.

7. Some modeling approaches permit the use of more direct proxies for *ex ante* risk. For example, Levonian, in an article in this issue of the *Economic Review*, calculates values of *ex ante* risk of bank stocks that are implied by the prices of option contracts on those stocks.

8. I also considered weighting the individual stock returns by the assets of the firms included in the groupings. This weighting did not significantly alter the results presented in Table 1.

9. I conducted Chow tests of the constancy of the set of estimated regression coefficients in the various subintervals of the sample period. For the two halves of the 12-year period, the F-value was 1.24. The critical value for the F-distribution, at a 99 percent confidence level and with (500, 1000) degrees of freedom, is 1.19. My half-sample test had approximately 6000 degrees of freedom in both numerator and denominator. Thus, the set of estimated coefficients in the two half-intervals were significantly different from one another. On the quarter-interval estimates, it was not possible to distinguish the first two quarters of the sample period: the F-value was 1.12, with approximately (3000, 3000) degrees of freedom. The third and fourth quarters of the sample were significantly different from one another: the F-value for this test was 1.46.

10. Chow tests on the constancy of the set of estimated coefficients in the two-factor model confirm that these coefficients changed significantly during the sample. The F-value between the two half-intervals was 1.36; between the first two quarters of the sample, 1.23; and between the last two quarters, 1.48. The critical value of the F-statistic at the 99 percent level, with (500,1000) degrees of freedom, is 1.19.

11. As mentioned in footnote 5, there are alternative ways to estimate cross-sectional differences in risk sensitivity. For example, size measures could be included as additional explanatory variables in the regression equations, or as variables interacted with the two risk factors. The method used here was chosen to highlight differences between banks in the different size categories.

REFERENCES

Beebe, Jack H. 1983. "Bank Capital Risk in the Post-1979 Monetary and Regulatory Environment," Federal Reserve Bank of San Francisco *Economic Review* (Summer) pp. 7-18.

________. 1985. "Bank Stock Performance Since the 1970s." Federal Reserve Bank of San Francisco *Economic Review* (Winter) pp. 5-18.

Booth, James, and Dennis Officer. 1985. "Expectations, Interest Rates, and Commercial Bank Stocks," *Journal of Financial Research* 8, pp. 51-58.

Campbell, Tim, J. Kimball Dietrich, and Mark Weinstein. 1985. "Some Evidence on Bank Holding Company Regulation: The Question of Expansion into the Insurance Business." In *Conference on Bank Structure and Competition*, pp. 587-616. Federal Reserve Bank of Chicago.

Chance, Don, and William Lane. 1980 "A Re-examination of Interest Rate Sensitivity in the Common Stocks of Financial Institutions." *Journal of Financial Research* 3, pp. 49-55.

Chen, Nai-fu, Richard Roll, and Stephen Ross. 1986. "Economic Forces and the Stock Market." *Journal of Business* 59, pp. 383-403.

Dietrich, J. Kimball. 1986. "Bank Stock Return Sensitivity to Market and Term Structure Risk." Unpublished manuscript. University of Southern California.

Fama, Eugene. 1976. *Foundations of Finance: Portfolio Decisions and Securities Prices*. New York: Basic Books.

Flannery, Mark, and Christopher James. 1982. "The Impact of Market Interest Rates On Intermediary Stock Prices." In *Conference on Bank Structure and Competition*, pp. 520-538. Federal Reserve Bank of Chicago.

________, and ________. 1984a. "The Effect of Interest Rate Changes on the Common Stock Returns of Financial Institutions." *Journal of Finance* 39, pp. 1141-1153.

________, and ________. 1984b. "Market Evidence on the Effective Maturity of Bank Assets and Liabilities." *Journal of Money, Credit and Banking* 16, pp. 435-445.

French, Kenneth, Richard Ruback, and G. William Schwert. 1983. "Effects of Nominal Contracting on Stock Returns." *Journal of Political Economy* 91, pp. 70-96.

Giliberto, Michael. 1985. "Interest Rate Sensitivity in the Common Stocks of Financial Intermediaries: A Methodological Note." *Journal of Financial and Quantitative Analysis* 20, pp. 123-126.

Kane, Edward, and Haluk Unal. 1988. "Change in Market Assessment of Deposit-Institution Riskiness." *Journal of Financial Services Research* 1, pp. 207-229.

Kwan, Simon. 1991. "Re-examination of Interest Rate Sensitivity of Commercial Bank Stock Returns Using a Random Coefficient Model." *Journal of Financial Services Research* 5, pp. 61-76.

Levonian, Mark. 1991. "Have Large Banks Become Riskier? Recent Evidence from Option Markets." Federal Reserve Bank of San Francisco *Economic Review* (Fall) pp. 3-17.

Lintner, John. 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics* 47, pp. 13-37.

Lloyd, William, and Richard Schick. 1977. "A Test of Stone's Two-Index Model of Returns." *Journal of Financial and Quantitative Analysis* 12, pp. 363-368.

Lynge, Morgan, and J. Kenton Zumwalt. 1980. "An Empirical Study of the Interest Rate Sensitivity of Commercial Bank Returns: A Multi-Index Approach." *Journal of Financial and Quantitative Analysis* 15, pp. 731-742.

Martin, John, and Arthur Keown. 1977. "Interest Rate Sensitivity and Portfolio Risk." *Journal of Financial and Quantitative Analysis* 12, pp. 181-195.

Merton, Robert. 1973. "An Intertemporal Capital Asset Pricing Model." *Econometrica* 41, pp. 867-887.

Rosenberg, Barr, and Philip Perry. 1981. "The Fundamental Determinants of Risk in Banking." In *Risk and Capital Adequacy in Commercial Banks*, ed. Sherman Maisel, pp. 367-407. Chicago: University of Chicago Press.

Ross, Stephen. 1976. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 13, pp. 341-360.

Sharpe, William. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance* 19, pp. 425-442.

Stone, Bernell. 1974. "Systematic Interest Rate Risk in a Two-Index Model of Returns." *Journal of Financial and Quantitative Analysis* 9, pp. 709-721.

Recession Probability Indexes: A Survey

Chan G. Huh

I wish to thank the editorial committee, Brian Motley, Adrian Throop, and Bharat Trehan, as well as John Judd for their helpful comments and Conrad Gann for his valuable research assistance. I also wish to thank James H. Stock and Glenn D. Rudebusch for providing data.

Specialized econometric models are designed to measure the likelihood of the occurrence of a recession in the near future. This paper examines a selected group of models that are distinct in terms of their theoretical underpinnings and also in terms of the scope of variables included. The models' performance of predicting the onset of the 1990 recession is mixed. In this case, it appears that what distinguished the models was less the difference in their theoretical underpinnings than whether or not the models included financial variables.

A wide range of methods is used to forecast recessions. For example, one method is a rule-of-thumb that predicts a recession following three consecutive declines in the Department of Commerce's composite index of leading indicators. At the other extreme are more advanced econometric models. This article will focus on the latter group, and in particular on two advanced econometric models that represent different theoretical approaches: the experimental recession probability index (XRI) model developed by James Stock and Mark Watson at the National Bureau of Economic Research (NBER), and a turning point forecasting model which implements a methodology proposed by Salih Neftci.

The experimental NBER model is theoretically similar to conventional linear regression models that are used for forecasting in general, although it is also unique in terms of the way information is extracted from data and in the information the data provide. Implicit in the NBER's XRI model and in other linear forecasting models is the assumption that expansions and contractions are part of the same stable structure, and that they are responses to random shocks (policy and otherwise).

Neftci-type turning point (TP) models depart from this key assumption of a stable structure. TP models posit multiple behavioral regimes that govern the movements of output over time. Thus, the process that best describes the behavior of output in an expansionary period is fundamentally different from the process describing the behavior of output in a contractionary period. Consequently, forecasting a downturn is equivalent to predicting a switch in the behavioral regime from an expansion to a contraction.

The recession of late 1990 provided the first out-of-sample opportunity to apply these models. The performances of the models in identifying this particular downturn as of late 1990 were mixed, giving probabilities ranging from 14 percent to 98 percent. Interestingly, the differences among the forecasts do not appear to be related to differences in their theoretical underpinnings, but rather to the types of variables used as the signaling source series. The models incorporating several financial variables are associated with low probability forecasts, and the models that rely mostly on nonfinancial variables result in high probability forecasts; that is, financial markets were, in this case, poor forecasters of the recession.

This result is not definitive, however, and must be placed in perspective. By definition, a forecasting model of stochastic outcomes cannot be expected to have a perfect fit repeatedly. One observation hardly provides enough information to judge the overall usefulness of the models. Or, as Aristotle, the father of logic, put it, "One swallow does not a summer make." To judge the accuracy, and thus the reliability, of these models requires a whole series of forecasts.

Brief overviews of different models for estimating the probability of economic downturns are provided in the next three sections: a standard linear regression model (Section I), the NBER's XRI (Section II), and the Neftci turning point forecast models (Section III). An overall assessment of their past within-sample forecasting performances is provided in Section IV, and a discussion of out of sample performance is presented in Section V. Section VI concludes.

I. FRB San Francisco BVAR Model: A Conventional Regression Model

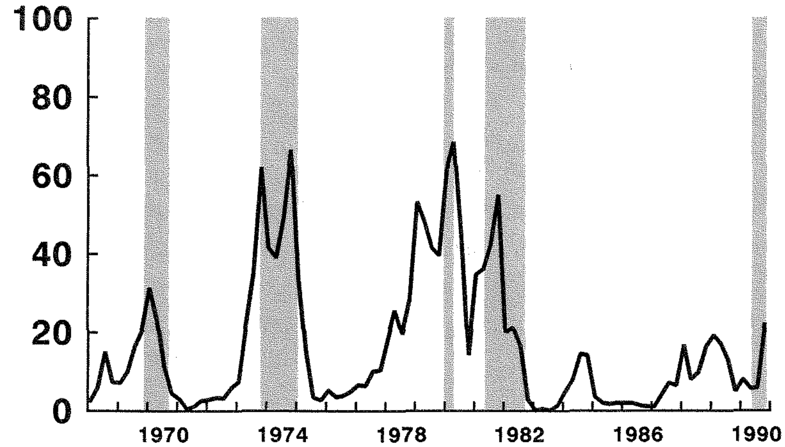
One easy and straightforward way to forecast a recession is to use linear regression models designed to forecast key macroeconomic variables. Any such regression model can be used to forecast a recession once the "operational" definition of a recession is determined in terms of variables in the model. One example is the Bayesian Vector Autoregression (BVAR) model used as a part of the FRB San Francisco in-house staff forecast.[1] The BVAR is designed to forecast growth in real GNP, inflation, and other key macro variables several quarters ahead.[2] The BVAR model is specified in terms of a combination of log and level differences of three real and seven nominal and financial quarterly variables.[3] The Bayesian prior affects this otherwise ordinary VAR system in the form of a priori restrictions on the magnitudes of the coefficients. For example, a simple prior restriction that the real GNP growth rate follows a random walk is imposed on the real GNP equation. Of course, the final estimates would reflect both this prior restriction and the sample information.

Suppose we are interested in finding the probability of a recession occurring within the next three quarters, and that we define a recession as two or more consecutive quarters of negative real GNP growth.[4] The following formula provides the necessary information to calculate the probability in period t:

(1) Prob (recession within 3 quarters) = Prob {event (output contracts in periods $t+1$ and $t+2$) $\cup$ event (output contracts in periods $t+2$ and $t+3$) $\cup$ event (output contracts in periods $t+1$, $t+2$, and $t+3$)}.

The actual calculation is in four steps. First, simulate a large number of unconditional forecasts (e.g., 1,000) for the next three quarters based on the model by repeatedly drawing from the stochastic error terms of the system. Second, count the number of simulated forecast triplets that fit any one of the three disjoint events that were defined in (1). Third, calculate the probability measure by dividing the sample numbers obtained from the second step by the total number of simulations. Fourth, the total probability of (1) is the sum of the three probability measures derived in the second step. Actual probabilities calculated this way from the FRBSF BVAR model are presented in Chart 1.

Chart 1
BVAR Recession Index



II. The NBER's Experimental Recession Index

The experimental NBER models are also based on the traditional regression method, and thus they share the key assumption that output series over time can be described by a single process. However, the experimental NBER models (Stock and Watson 1989) are more specialized in terms of their scope and of the econometric technique employed.

The NBER XRI is based on two artificial signaling index variables that, in turn, are constructed from sets of actual economic variables. The signaling variables are the experimental indexes of coincident economic indicators and of leading economic indicators.[5]

The experimental index of coincident economic indicators (CEI) is designed to measure the level of current economic activity. It involves a weighted average of four series that are widely perceived to be coincidental: industrial production, real personal income less transfer payments, real manufacturing and trade sales, and employee-hours in nonagricultural establishments. The index is based on a dynamic factor model that measures the change in an unobserved factor that is assumed to be a significant source of movement in all four series (for details, see Sargent and Sims 1977). In terms of both cyclical behavior and historical trend, Stock and Watson's CEI is very close to the CEI released by the Commerce Department, which also was designed to reflect the general state of the economy. The main differences between the two are that Stock and Watson use newer econometric technology to construct the overall index from its components and that Stock and Watson use the employee-hours series, while the Commerce department uses the number of employees.

The experimental index of leading economic indicators (LEI) was designed to provide optimal forecasts of the projected growth in the CEI over the next six months given the information up to period t. There are two versions of the LEI, namely, the XLI and the XLI-2, which differ in the variables they use. The XLI uses seven variables—three real and four nominal and financial variables—that were selected after applying multiple sets of tests to a large number of candidate variables. The XLI-2 replaces all nominal and financial variables used in the XLI, except the

Table 1
List of Variables

NBER Experimental Leading Index		DOC Leading Indicator
XLI	XLI-2	
1. New private housing authorized index	1. New private housing authorized index	1. Average weekly hours, manufacturing
2. Manufacturers' unfilled orders: Durable Goods Industries	2. Manufacturers' unfilled orders: Durable Goods Industries	2. Average weekly initial claims for state unemployment insurance
3. Trade-weighted dollar	3. Trade-weighted dollar	3. Manufacturer's new order for consumer goods and materials industries
4. Part-time work in non-agricultural industries because of slack work	4. Part-time work in non-agricultural industries because of slack work	4. S&P 500
5. Yield on constant-maturity portfolio of 10-year Treasury Bonds	5. Help wanted advertisement	5. Contracts and orders for plant and equipment
6. Spread between interest rate on 6-month CP and 6-month Treasury Bonds	6. Capacity utilization rate	6. New private housing authorized index
7. Spread between variable 5 and the yield on 1-year	7. Average weekly work hours, manufacturing	7. Vendor performance, slow deliveries diffusion index
	8. Vendor performance, slow deliveries diffusion index	8. Changes in sensitive materials prices
		9. Money supply M2
		10. Change in Manufacturers' unfilled orders, durable goods industries
		11. Index of consumer expectations

exchange rate, with additional real variables. (See Table 1 for the list of variables. For a detailed econometric description, see Stock and Watson (1989) or Watson (1991, Appendix).)

Two related experimental recession indexes are based on these models, the XRI and the XRI-2. These indexes are designed to measure the probability that the economy (gauged by the CEI) will be in a recession six months hence.[6] Actual probabilities using a stochastic simulation method that is similar to the procedure described in Section I for XRI and XRI-2 are presented in Charts 2 and 3, respectively.

This procedure is valid under a key assumption of linearity in the relationship between the variables involved. That is, the underlying model that describes the behavioral relationship between real GNP (or any variable of interest) and its explanatory variables must remain stable and symmetric across both expansionary and contractionary phases of business cycles. If this linear relationship does not hold, one has to consider some alternative ways to describe the behavioral relationship. Subsequently, calculating the probabilities of the events defined in (1) would become more involved.

Indeed, some economists think that there are fundamental differences in the behavioral patterns of key variables across expansion phases and contraction phases of business cycles. We now turn to a specialized recession probability model that is based on such a view.

Chart 2
NBER Recession Index (XRI)

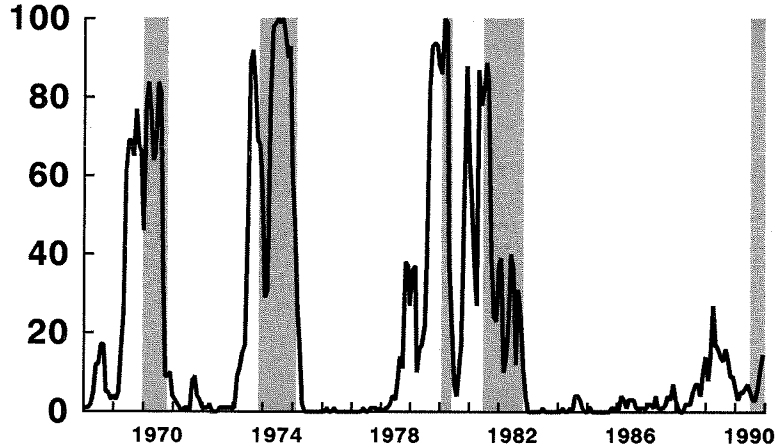
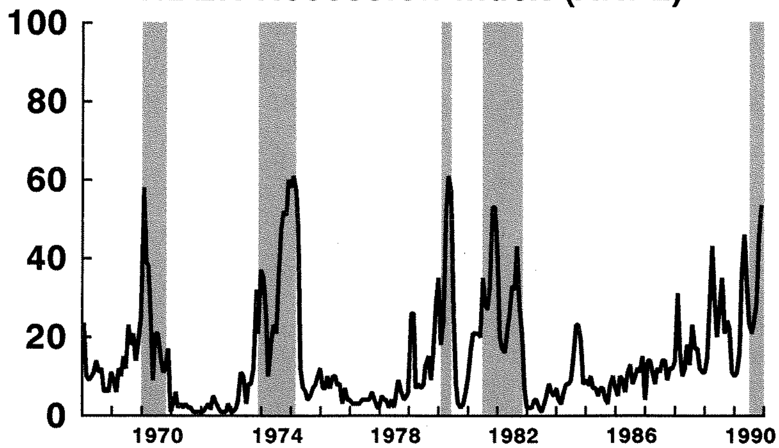


Chart 3
NBER Recession Index (XRI-2)



III. Turning Point Recession Index: Process Switching Model

Many economists have observed asymmetric behavior of some key macro variables between economic expansions and contractions. For example, output tends to inch upward during an expansion, but it tends to drop very sharply at the beginning of a contraction. Thus, the behavior of the economy in the two phases is best described as being governed by two distinct stochastic structures instead of by a single underlying structure (Neftci 1982). Consequently, according to these views, forecasting a recession (the onset of a contraction regime) amounts to predicting a behavioral switch in the economy from an expansion to a contraction.

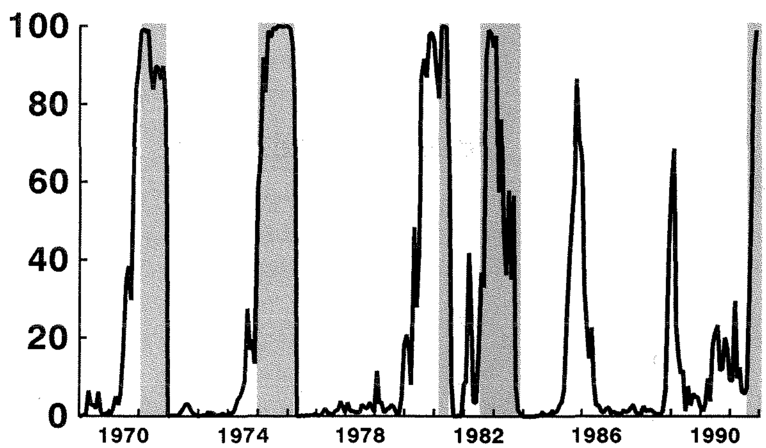
To put this idea into practice, a forecaster needs a signal variable that foretells changes in the behavioral structures. This signal variable must meet several requirements: its behavior should be systematically related to that of the output in the economy; it should have some lead time with respect to changes in output to be useful as a predictor of changes in output; finally, it should be available frequently enough to update the model in a timely manner in terms of key developments that have a bearing on the potential shift in the regime. Both the original turning point model of Neftci (1982) and a model by Diebold and Rudebusch (1989) use the monthly Composite Index of Leading Indicators published by the Department of Commerce (henceforth, DOC LI) as such a signal variable.

The next step is to take the first difference of the series. Then the data characteristics in upturns and downturns are summarized by fitting simple normal distribution functions: first divide the overall historical period into expansionary and contractionary sub-periods using the historical turning point dating in the DOC LI series. Then the DOC LI observations belonging to expansionary and contractionary periods are respectively pooled into two groups of upturn and downturn samples. Finally, two normal distribution functions $N^e(\mu_e,\sigma_e)$, $N^c(\mu_c,\sigma_c)$ (where μ and σ denote the mean and standard deviation) are estimated from the expansion and contraction samples, respectively.

Additionally, determine a prior transitional probability for the signal variable. The transitional probability Ω is the measure of the likelihood that the signal variable will remain in the current regime at any given time.[7] Consequently, $1-\Omega$ measures the probability of the signal variable switching from the current behavioral regime.

Given this information, one can apply the switching time Bayesian probability formula that was developed by Neftci (1982) as shown in Box 1. With each new observation in the DOC LI, the turning point recession index (TPRI) model calculates the conditional probability that the indicator is in the downturn regime. The probabilities are shown in Chart 4.

Chart 4
Turning Point Recession Index



IV. Assessment

Recession probability indexes and turning point models are more sophisticated than rule-of-thumb methods, which typically forecast a recession after three consecutive declines in the DOC LI. They are also more systematic because they account for the magnitude of change as well as the temporal direction of change in the leading indicators, and in general, they substantially outperform the rule-of-thumb predictions.

Stock and Watson (1989, pp. 382) applied a formal econometric method to compare the predictive power of

Box 1

Denote the signal variable (i.e., DOC LI) at date t as i_t and the date when a switch occurs from one behavioral regime of i_t to another as Z. Z will be a random variable that takes integer values. Next denote the collection of the sequential realizations of i up to the present (time t) as I_t. Then the posterior probability for a turning point occurring at time t, given the data I_t can be written as follows:

$$P(Z \leq t | I_t) = \frac{P(I_t | Z \leq t)}{P(I_t)};$$

or, by denoting $P(Z \leq t | I_t)$ as Π_t, we get

$$\Pi_t = \frac{[\Pi_{t-1} + \Omega(1-\Pi_{t-1})]p_t^d}{[\Pi_{t-1} + \Omega(1-\Pi_{t-1})]p_t^d + (1-\Pi_{t-1})p_t^u(1-\Omega)},$$

where: Π_t denotes the conditional probability that the DOC LI is in the downturn regime in period t;

p_t^u, p_t^d denote the probabilities, obtained from the step described earlier, that the observed i_t came from $N^u(\mu_u, \sigma_u)$ (upturn regime) and $N^d(\mu_d, \sigma_d)$(downturn regime), respectively;

Ω denotes the unconditional transitional probability of a switch from an upturn regime to a downturn regime.

A new observation on i_t will affect the conditional probability of Π through p_t^u and p_t^d because other elements are predetermined. For example, suppose that currently we observe a large positive value for i_t. Then it is more likely that it came from the expansionary regime than from the contractionary regime, because on average, the economy grows more during expansions than contractions; that is, $p_t^u > p_t^d$. This large p_t^u will make the second part of the denominator larger relative to both the numerator and the first entry of the denominator. It makes the overall Π_t smaller, and thus makes a switch of regimes (from upturn to downturn) in the current period less likely. This consequently reduces the probability that the economy will be in recession in the near future. This formulation works symmetrically for a forecast of a regime switch from downturn to upturn with the proper substitutions of terms.

the rule-of-thumb method and their method, and found their method to be more accurate. They performed regression analyses that related consecutive movements in the DOC LI numbers to actual historical recessions and expansions. For example, the R^2 of the regression that related the index to the occurrence of recessions or expansions six months hence was 0.028 using the rule-of-thumb, whereas it was 0.50 for the regression using the experimental NBER XRI. Diebold and Rudebusch (1989) also found a similar relative performance ranking of the rule-of-thumb method compared to different methods, such as the Neftci method.[8]

However, assessing the "goodness of fit" of these models is conceptually difficult, because their forecasts are in terms of probabilities, and "actual" probabilities are not directly observable to evaluate the performance of these models. In fact, low probability recessions may occasionally occur, while high probability recessions may occasionally not occur. Thus, over a limited sample period, simply correlating the probability with the business cycle is not necessarily a good way to judge the accuracy of the models. Of course, the larger the sample, the more appropriate this direct type of evaluation becomes.[9]

One criterion that is often used to gauge reliability is the frequency of false signals, a notorious problem for leading indicators that led to Paul Samuelson's famous remark, "The stock market has predicted nine out of the last five recessions!" The first type of false signal is analogous to the Type II error of the usual hypothesis test; that is, the model forecast of an imminent recession is not followed by an actual recession within a reasonable period of time. For example, suppose we interpret a model as signaling a recession when the probability is above half of the maximum probability (observed over the sample period) of each model. According to this criterion, the TPRI and NBER XRI-2 models each have two instances of false signals for the 1968-1989 sample period (1985 and 1988 for TPRI and 1988 and 1989 for XRI-2); the very striking spikes in the TPRI model in the late 1980s seem to have been reflecting temporary slowdowns in the manufacturing sector during those periods. The NBER XRI and BVAR RI have no false signals.

The second type of false signal is analogous to the Type I error of the usual hypothesis test; that is, the model fails to predict an ensuing recession with some lead time (for example, six months). Using the same cut-off probability as in the first case, the NBER XRI-2 has failed four times (1969, 1973, 1980 and 1982), the TPRI model has failed twice (1974 and 1981), the BVAR RI has failed once (1969), and the NBER XRI has not failed at all. The performance of the TPRI in this regard is most likely related to the fact that it is based on the DOC LI which is notorious for having widely varying lead times with the business cycle peaks and troughs. For the past 30 years, for example, turning points in the DOC LI have led the contractionary turning points of the economy by anywhere from two to twenty months.[10]

VI. Recent Predictions

As in any of the forecasting models that have been estimated using sample information, an important test of the RPI models' forecasting power hinges on their out-of-sample performance. The only out-of-sample observation we have is the most recent recession, which started in the second half of 1990.

The models' various predictions of the probability of an imminent recession as of the end of November 1990 are 14 percent for the NBER XRI, 21 percent for the BVAR, 53 percent for the NBER XRI-2, and 98 percent for the TPRI. The sharp divergence between the forecasts of the NBER XRI and XRI-2 suggests that different theoretical underpinnings alone do not explain the divergent forecasts. It is natural to ask, then, what the most likely source of such differences is.

One distinguishing feature is that low probability forecasts included a set of financial variables (interest rates and associated spreads) but high probability forecasts did not. This is particularly interesting in light of recent studies on the changing role of financial variables in econometric models of key macro variables.

Bernanke (1990), among others, found that various interest rates and spreads were substantially more useful in explaining and forecasting key macro variables for the pre-1980 sample period than for the post-1980 period. In particular, he examined the spread between the commercial paper rate and the T-bill rate. This spread may reflect the default risk of commercial paper, which, in turn, would be very sensitive to an expected recession. However, if this is the important channel through which the financial variables are useful in forecasting key variables, then they still can be expected to have substantial explanatory power in econometric models.

The spread may also reflect the monetary policy stance, which affects the near-term economic condition by shifting credit conditions. This may be particularly relevant when there are deposit interest rate ceilings and when commercial paper and T-bills are imperfect substitutes as portfolio assets. Monetary tightening would induce an outflow of deposits from banks as market interest rates rise above deposit ceilings. This "disintermediation" creates a "credit crunch" and subsequently an economic contraction. At the same time, bank deposits will flow into T-bills because T-bills can be purchased in relatively small denominations, unlike commercial paper, which is typically in denominations too large for most small deposit holders. This inflow of funds will depress T-bill yields relative to commercial paper rates in periods when the general level of interest rates is higher.[11]

According to this hypothesis, it is relatively easy to explain the diminished role of the spread. Since the early 1980s when deposit rates were deregulated, more alternative financial assets have become available creating closer substitutability among assets.

This conjecture seems relevant in explaining the divergent forecasts of the various models. The yield curve has maintained a positive slope, and few noticeable changes with respect to short-term interest rates and the rate-spread have occurred in late 1990. Thus, the recession forecasts of models that included these financial variables might have picked up mixed signals of the likely conditions of the economy, unlike the models with only real variables. Consequently, according to this conjecture, the probability of recession forecasted by models containing financial variables did not increase as substantially as it did in models relying entirely on real variables.

It is quite possible that the current recession is distinct from preceding ones in terms of both its causes and the way contractionary effects of the causal factors spread across the economy. For example, some economists cite the diminished credit availability which started in 1990 for reasons related to the weakened condition of financial institutions and stricter regulations while others point to the special circumstances associated with the Middle East confrontation, which increased short-term and near-term uncertainties.

The question of whether business cycles are distinct (and hence whether a single modeling strategy is appropriate) is not new. Blanchard and Watson (1986) examined the nature of the sources of impulses behind business cycles using U.S. time series data. Their findings suggest that cycles are not alike; that is, each historical cycle can be associated with several identifiable large single shocks with different origins.

This result, however, does not necessarily make the modeling approaches surveyed here inappropriate. The existing models are still valid and applicable to the designed task if there exist measurable similarities in the way the original shocks propagate or dissipate throughout the economy. Whether or not this is the case is an important empirical issue, one on which we can expect to see more research in the future.

VII. Conclusion

Representative models designed to forecast prospects of a recession in the near future have been examined. Specifically reviewed were the experimental NBER index models and a model based on the Neftci method. They differ not only in various operational aspects, but also in their conceptual approaches to modeling the behavior of key economic variables, such as output. The experimental NBER models are based on the assumption that output behaves symmetrically across both expansionary and contractionary phases of economic fluctuations, whereas the Neftci method admits a shift in the behavioral regime across the two phases. Whether this assumption of a symmetry in the behavior of the output is empirically appropriate is an issue currently being examined by economists.

The models performed well in terms of within-sample historical predictions. They outperformed the common rule-of-thumb that relies on three consecutive declines in the DOC LI. However, their out-of-sample forecasts were widely divergent, even for those that used the same modeling approach. The distinction is that models with a high probability forecast excluded a set of financial variables while low probability forecasts included financial variables. This seems to reflect the fact that a recession is defined to be a period of contractions in real variables such as orders, sales, output and employment. Although the amount of the lead time may vary, models that rely comprehensively on such real variables will necessarily provide indications of the onset of a serious downturn.

It is likely that the most recent downturn was unusual in that its causal factors differed from the few factors that had frequently been behind past recessions. In that sense, the models that were designed to conform to the general average characteristics of past economic fluctuations did poorly in detecting the most recent economic downturn.

However, it is premature to draw any inferences from this single-sample observation of the current recession; these results need to be considered in the proper perspective. In most situations where we need to draw inferences about an uncertain outcome, more information is preferred to less as a practical principle. This holds true with regard to forecasting business cycle downturns, especially since we do not have a well-understood, widely agreed upon, and operationally feasible framework for describing evolutions of a large set of macro variables.

Such a framework could provide a theoretically well-founded list of variables or a sequence of economic events that could give rise to a "sufficient statistic" about a near-term economic downturn, and would consequently make any additional information redundant. In this context, systematic efforts to reduce our prediction errors involving important aggregate economic variables such as the RPI models can be useful.

The key contribution of the RPI models, however, essentially lies in providing another way to organize and use information contained in the various leading economic indicators. Consequently, their reliability is crucially dependent on the reliability of the leading economic indicators that are used as the sources of information. Thus, any further refinement and improvement of our stock of knowledge on leading indicators will lead to commensurate improvement in the performance of the recession forecasting models.

ENDNOTES

1. For detailed explanations of this modeling strategy, see Todd (1984) and Roberds (1988). For a more theoretical discussion of this econometric methodology see Doan, Litterman and Sims (1984).

2. This is done by weighing forecast accuracy at the one-year horizon more heavily than the rest of the forecasting horizon. This step is implemented during a model specification selection stage. Following the BVAR modeling practice, a model builder defines a prior matrix of parameters that control the dynamic interactions between variables in each equation of the model. We adjusted these prior parameters selectively to obtain the forecast accuracy configuration across different forecasting time frames. See Roberds (1988) for detailed descriptions.

3. The variables are real GNP, business fixed investment, the unemployment rate, fixed GNP price index, unit labor cost, producer price index, monetary aggregate M2, trade weighted exchange rate, six-month commercial paper, and AAA corporate rates.

4. Applying this rule to post-war U.S. data (1947Q1-1990Q3) we detect six out of the eight recessions that occurred.

5. These indexes were developed as the result of efforts to update the system of indicators that were developed in the 1930s and 1940s at the NBER by Mitchell and others; the latter is still being used at the Department of Commerce.

6. This particular time horizon is related to the way the LEI is constructed. That is, it was specifically designed to give an optimal forecast of the CEI's relative growth over *six months*. A more precise definition of the economy being in a recession is as follows: A month is defined to be in a recession pattern if the monthly growth of the CEI index is either in a sequence of six consecutive declines below a boundary point, or in a sequence of nine declines below the boundary with no more than one increase during the middle seven months.

7. There are different views regarding the question of dependency between the duration of each phase (i.e., expansionary or contractionary) and the probability of transition from one regime to another. For example, in Neftci (1982), the transition probability is treated as dependent on the duration, whereas Diebold and Rudebusch (1989) use a transitional probability matrix that is independent of the duration. For more details on this issue, see Hamilton (1989), Neftci (1984), Diebold and Rudebusch (1990).

8. However, a recent study by Koenig and Emery (1991) found that some relatively simple methods similar to the rule of thumb did as well as the Neftci method when the actual *real time* data that would historically have been available to a forecaster were used, instead of the most recent revised data on the DOC LI. These results point to some potential problems with the DOC LI series, which has gone through major revisions, rather than to a devaluation of the Neftci methodology.

9. Diebold and Rudebusch (1989) propose and examine a set of test statistics that can score a probability forecast model in terms of different attributes such as accuracy, calibration and resolution. Even though these proposed methods are systematic, small-sample observations are still problematic. However, the turning point forecast model seems more appropriate for such an evaluation method, because it generates more observations both in terms of switches from expansions to contractions and vice versa, whereas a simple recession forecast model would count only switches from expansionary to contractionary regimes.

10. Koenig and Emery (1991) give a detail account of the relative performance of the real time DOC LI series in predicting expansions versus contractions. They find the series to be a better predictor of expansions and a poorer predictor of recessions in near future.

11. Financial institutions would have an opportunity to arbitrage by selling T-bills in their portfolio and buying commercial paper in those periods. However, for banks those two instruments are not alike. For example, banks can use T-bills but not commercial paper as collateral for satisfying bank capital adequacy. Thus, due to the imperfect substitutability between T-bills and commercial paper, banks will not arbitrage and offset the widening spread.

REFERENCES

Bernanke, Ben S. 1990. "On the Predictive Power of Interest Rates and Interest Rate Spreads," Federal Reserve Bank of Boston *New England Economic Review* pp. 51-68.

Blanchard, Olivier J., and Mark W. Watson. 1986. "Are Business Cycles All Alike?" In *The American Business Cycles*, ed. R.J. Gordon. Chicago. The University of Chicago Press.

Diebold, Francis X., and Glenn D. Rudebusch. 1989. "Scoring the Leading Indicators." *Journal of Business* 64, pp. 369-391.

_________. 1990. "A Nonparametric Investigation of Duration Dependence in the American Business Cycle." *Journal of Political Economy* 98, pp. 596-616.

Doan, Thomas, R. B. Litterman, and C. A. Sims. 1984. "Forecasting and Conditional Projection Using Realistic Prior Distribution." *Econometric Review* 3, pp. 1-100.

Hamilton, James D. 1989. "A New Approach to the Economic Analysis of Non-Stationary Time Series and the Business Cycle." *Econometrica* 57, pp. 357-384.

Koenig, Evan F., and Kenneth M. Emery. 1991. "Misleading Indicators? Using the Composite Leading Indicators to Predict Cyclical Turning Points." Federal Reserve Bank of Dallas *Economic Review* (July) pp. 1-14.

Neftci, Salih N. 1982. "Optimal Prediction of Cyclical Downturns." *Journal of Economic Dynamics and Control* 4, pp. 225-241.

_________. 1984. "Are Economic Time Series Asymmetric over the Business Cycle?" *Journal of Political Economy* 92, pp. 307-328.

Roberds, Will. 1988. "A Quarterly Bayesian VAR Model of the U.S. Economy." Working Paper 88-2. Federal Reserve Bank of Atlanta.

Sargent, T. J., and C. A. Sims. 1977. "Business Cycle Modeling without Pretending to Have Too Much A Priori Economic Theory." In *New Methods in Business Cycle Research*, ed. C.A. Sims, et al. Minneapolis: Federal Reserve Bank of Minneapolis.

Stock, James H., and Mark W. Watson. 1989. "New Indexes of Coincident and Leading Economic Indicators." *NBER Macroeconomics Annual* pp. 351-394.

Todd, Richard M. 1984. "Improving Economic Forecasting with Bayesian Vector Autoregression." Federal Reserve Bank of Minneapolis *Quarterly Review* (Fall) pp. 18-29.

Watson, Mark W. 1991. "Using Econometric Models to Predict Recessions." Federal Reserve Bank of Chicago *Economic Perspectives* (November/December) pp. 14-25.

The Use of Equity Positions by Banks: The Japanese Evidence

Sun Bae Kim

Economist, Economic Research Department, Federal Reserve Bank of San Francisco. This paper is partly based on a chapter of my dissertation submitted to the Department of Economics, University of Toronto. I would like to thank the Center for Japanese Studies, Graduate School of Business, Columbia University, for granting me access to the NEEDS financial tapes. I am particularly indebted to Profs. Hugh Patrick and Masako Darrough for kindly making such an arrangement. Grateful acknowledgement is also extended to Hang-Sheng Cheng, Mark Levonian, Ramon Moreno, Brian Motley, and Randy Pozdena, for their comments and suggestions. Able research assistance was provided by Sean Kelly.

This paper draws on Japan's experience to analyze the merits of conferring securities powers on banks. By issuing debt and equity jointly to a single outside investor, the shareholder bank, firms will incur lower deadweight costs associated with bankruptcy and monitoring than if these claims are issued to separate entities. The level of Japanese banks' shareholding in firms during the late 1960s is generally consistent with this explanation of financing decisions. Most notably, banks as a rule provided joint debt-equity financing rather than pure debt financing. Moreover, holding other things constant, the level of a bank's equity holding increased in proportion to financing it supplied the firm and to the riskiness of investment.

A broad consensus now prevails that a firm in need of external financing for investment may incur deadweight losses that another firm, with sufficient internal funds but otherwise identical, would not. The basic idea is that when a firm knows more than outside financiers about its own prospects and/or actions that affect its prospects, external financing will engender deadweight or "agency costs" which ultimately fall on the firm. Credit- and equity-rationing, the need to meet strict collateral and other financial requirements, and resources expended on monitoring, are some of the manifestations of such costs.

One important conclusion from the literature is that the more the corporate sector relies on external funds, the greater will be the potential magnitude of agency costs. In such a setting, the extent to which the market can devise and implement contracts which attenuate these information-related costs will have a significant bearing on the "real" performance of the economy. Financial contracts do not materialize in a vacuum, however. The extent to which agents can mitigate capital market imperfections is dictated by the legal and institutional parameters of the financial system itself. Notably, we employ financial contracts with constraints idiosyncratic to the Anglo-American system, namely, the separation of commercial and investment banking. There is no reason to believe, however, that such a systemic constraint should remain binding, much less be desirable, across time and space. If the corporate sector's dependence on external financing is sufficiently heightened, for instance, the constraint itself may be subject to change.

The postwar Japanese experience is a case in point. The rapid investment-led growth from the 1950s to the early 1970s put a formidable burden on Japan's financial system. By virtue of the pace of growth, industries' demand for external funds was large relative to their net worth or collateral and hence, the potential agency costs in issuing debt and equity were commensurately high. In such a setting, the banks, which were the primary conduit of investable funds, were legally sanctioned simultaneously to extend loans and to hold shares of client firms.[1] The predominant mode of financial contracts during Japan's rapid growth period thus featured the major lenders as also significant shareholders. Judging from the performance of its economy, such a system appears to have met admirably the task of underwriting Japan's growth.

Drawing on Japan's experience, this paper explores the possible merits of conferring on banks securities powers.[2] It does so in two steps. Section I motivates the relevance of financing choice on firm value under two types of capital markets imperfections: costly bankruptcy and imperfect information. Two points are established here. First, the value-maximizing financial contract will optimally trade off the agency cost of debt against that of equity. Second, total agency cost of external finance will be lower when debt and equity claims are jointly issued.

If agency costs do indeed matter in firms' financing decisions, then firm-specific parameters affecting the relative severity of the agency costs of debt versus those of equity should explain observed interfirm differences in ownership structure. Section II explores this issue by focusing on the determinants of the level of the bank's equity holding for a cross-section sample of Japanese firms during the period of rapid growth. The main contribution here is the attempt to ascertain the effects of firm-specific parameters on the level of the bank's equity claims that is *jointly* held with debt claims. This approach contrasts with previous studies that have focused on determinants of more "conventional" measure of firms' capital structure such as leverage ratio.[3]

I. An Agency Cost Approach to Corporate Finance[4]

Deadweight Costs of External Financing

Consider an entrepreneurial (i.e., owner-managed) firm facing an investment that requires external financing. Its return depends on a productive input we shall refer to as "effort," whose level is set by the owner-manager. Higher levels of effort enhance the return on the investment, but at a cost of increasing the disutility incurred by the manager. The investment is risky because it also depends on a random variable the manager does not control.[5]

The owner-manager can raise the required amount of external financing by issuing debt and/or new equity. Assume that banks are the only lenders in the system.[6] Furthermore, to serve as a benchmark, assume initially that banks cannot own equity; that is, lenders in the system cannot be shareholders. If the investment is undertaken, the firm can then be viewed as the nexus among three types of claimants: the original shareholder (the owner-manager), the new shareholder, and the lender (bank). As Modigliani and Miller (1958) demonstrated, the value of the firm will be independent of its financial structure if capital markets are perfect. I motivate the relevance of the financing decision by the owner-manager by positing two types of imperfections.

First, bankruptcy is assumed to be costly in the sense that the transfer of assets from the shareholders to the creditors consumes some fraction of their total value. These deadweight costs include lawyers' and accountants' fees, and costs incurred in auctioning off the firm's assets. Though more difficult to measure, bankruptcies also impose indirect costs which may turn out to be even larger.[7] Legal disputes or the perceived conflict of interest between the shareholders and creditors of the firm can distract and constrain the managers from properly running the business. Customers may desert financially distressed firms and suppliers may exact more favorable terms. If lenders are rational, these costs will be anticipated when negotiating the loans. Consequently, the expected deadweight cost of bankruptcy will ultimately fall on the borrower.

The second set of imperfections relates to asymmetric information between the firm and outside suppliers of funds, the creditor (bank) and the *new* shareholder. This gives rise to two types of incentive or moral hazard problems. Consider first the case when new shares are issued. As the owner-manager cedes larger proportions of the firm's residual profit to the new shareholder, the effort level, which is private information to the owner-manager, will decline.[8] This obtains from a standard assumption in principal-agent models: The outside shareholder (principal) shares in the fruits of the insider's (agent's) effort, but not in the level of effort itself. As a result, the agent provides less effort than the principal deems optimal. Insofar as the profitability of the firm depends on effort, the prospective shareholder will impute the effect of this adverse incentive when pricing the firm's equity. The inside shareholder will therefore suffer a reduction in the value of his equity.[9]

Deadweight costs will arise with issues of debt as well. First, the shareholder has an incentive to gain at the expense of the lender by increasing the riskiness of the investment the firm undertakes. This obtains because if a risky investment turns out to be very profitable, the shareholder captures most of the gain, while the maximum return to the lender is fixed to the contracted interest plus principal. If, on the other hand, the investment fails, limited liability will shield the shareholder and the lender will bear the consequences. This asymmetry in the payoff—bounded maximum downside loss but unbounded upside return—creates an incentive for the shareholder to invest in risky projects even if they have lower net present value than a safer project.[10]

The rational lender will anticipate the risk-taking propensity of the shareholder and demand an interest rate commensurate with the maximum amount of risk that the firm can undertake. In other words, by anticipating the full extent of the moral hazard risk-shifting by the shareholder, the lender will preempt any possible appropriation of his wealth. But such a preemptive measure will prove costly to the shareholder—and ultimately to society. Other things equal, a higher interest rate implies a higher probability of bankruptcy and hence, a higher expected cost of bankruptcy. In a competitive debt market, this higher expected cost will be passed on to the shareholder.

When the level of effort is private information to the insider, issues of fixed claims to outsiders will give rise to yet another deadweight cost. The level of effort set by the manager will decline as the probability of bankruptcy increases. This is a rational response because when the firm is bankrupt, its ownership will fall into the creditor's hands and the manager will not reap any return to expanding effort. Accordingly, effort will decline as interest rates rise, since the latter implies higher probability of bankruptcy. But reduced effort implies greater probability, and hence greater expected cost, of bankruptcy. Again, this deadweight cost will be passed onto the firm.

To summarize, the firm will incur deadweight losses whether it issues debt or equity. Substituting debt for equity substitutes one set of deadweight losses for another, but does not eliminate them, so long as the firm needs external financing. Under such circumstances, the firm's financing decision will matter because by optimally trading off the deadweight cost of debt against that of equity, total deadweight cost will be minimized. That firms do not exclusively rely on debt or on equity presumably reflects the fact that such "corner solutions" are not optimal from the point of view of minimizing agency costs.

Advantages of Joint Debt-Equity Financing

The joint issue of both fixed and residual claims to a single outside financier—the lender-cum-shareholder bank—can be motivated by two sets of considerations. The first is reductions in bankruptcy costs, direct as well as indirect. To the extent that the debtholder already owns a fraction of the firm, the total cost that will be incurred in transferring the firm's asset from the shareholders to the creditor in the event of bankruptcy will be smaller than when these claim holders are mutually exclusive entities.

The second set of advantages concerns the efficiency gains in monitoring. Up to this point, it has been assumed that outside financiers anticipate the full extent of moral hazard and price the firm's securities accordingly. However, the severity of the moral hazard problem, and hence its deadweight costs, can be attenuated by monitoring the activities of the insider. It will be ultimately in the interest of the owner-manager to subject himself to such scrutiny if the total monitoring cost is less than the reduction deadweight costs that it brings about, since the insider ultimately bears the agency costs in either case.

A commonly discussed pattern in the literature is that the lender checks the borrower's propensity to undertake risk through monitoring.[11] The outside shareholder monitors the effort of the manager, who acts in the interest of inside shareholders, and thereby enhances the net return to the investment. As in other forms of information production, however, monitoring exhibits economies of scope. Needless duplication of monitoring costs will be eliminated by having a single outside financier, i.e., the lender-shareholder, performing the monitoring. The idea can be illustrated with a simple example.

Suppose that the outside shareholder incurs X units of monitoring cost to increase the effort level of the manager by Δe. Analogously, suppose that the lender also devotes X units of resource to reduce the riskiness of the firm by $\Delta\sigma^2$. This will have two effects. First, the reduction in risk will lower the bankruptcy probability and hence the expected cost of bankruptcy. Second, the lowered bankruptcy probability, in turn, will raise the marginal expected return to effort for the owner-manager, and thus will elicit higher effort. Thus, the shareholder's monitoring of effort, which is costly, will be partly redundant as long as the lender's monitoring of risk has some positive spill-over effect on effort. By the same logic, the lender's monitoring of risk will be redundant, in part or in whole, when the shareholder is monitoring effort, since higher effort lowers bankruptcy probability and hence its expected cost.

Why can't the lender and shareholder coordinate the task of monitoring so as to eliminate any duplication? One obvious obstacle to such coordination is the agency conflict that prevails between the two. For example, while the lender can safely rely on the incentive of the shareholder to monitor the effort of the manager, he cannot trust the shareholder to monitor risk with commensurate self-interest; the shareholder prefers higher to lower risk. So long as a dissonance of interest prevails between the lender and the shareholder, the bundling of debt and equity claims into a single entity (i.e., the lender-shareholder bank) is a more efficient way to capture the economies of scope in monitoring, and ultimately to reduce the deadweight costs of external financing.

Hypotheses

While agency and bankruptcy costs provide a theoretical justification for the relevance of financing decision, are these costs indeed significant in reality? This question can be addressed more formally through a number of hypotheses suggested by the theoretical framework. The most obvious hypothesis is: if issues of debt and equity do impose deadweight losses, the prevalent mode of financing should feature the lending bank as a shareholder in the same firm, provided, of course, that the legal system permits it. In addition to this broadly cast prediction, more specific hypotheses emerge on interfirm differences in the level of equity stake held by lenders.[12]

HYPOTHESIS 1: the bank's shareholding will be larger the higher is the magnitude of deadweight losses holding the probability of bankruptcy constant. As the creditor holds larger equity stake in the firm, we would expect these deadweight costs to decline on sheer logistic grounds: a smaller portion of the firm's assets need to change hands. More significantly, perhaps, the greater coincidence of interest between debt and equity that obtains by definition in a lender-shareholder financing scheme will reduce the indirect cost of bankruptcy, or even reduce the probability that the firm encounters financial distress in the first place.

HYPOTHESIS 2: The higher is the firm's dependence on the bank's funds, the higher will be the bank's equity stake in the firm. All other things equal, the more the firm relies on outside financing, the higher will be the agency costs of debt as well as that of equity.[13] Greater equity holding by the bank may attenuate these deadweight costs on two grounds. First, by structuring a greater portion of the bank's return through equity, the expected costs of bankruptcy will be reduced. Second, increased shareholding may enhance the efficiency as well the clout of the bank in checking both types of moral hazards discussed above, i.e., risk-shifting and shirking on effort.

HYPOTHESIS 3: Holding the level of risk constant, the higher the expected profitability of the firm, the lower should be the bank's equity holding in the firm. Higher profitability implies lower probability of bankruptcy and hence lower expected deadweight cost of debt.[14] On the other hand, higher profitability will exacerbate the incentive problem associated with issues of (outside) equity: Outside financiers share an enlarged pie while the original owner-manager incurs all the cost of making it. The two effects combined should lead to a lower reliance on equity.

HYPOTHESIS 4: the bank's shareholding will be higher in riskier firms. Holding expected return constant, higher risk implies higher probability of bankruptcy and hence higher expected cost of bankruptcy. By structuring larger portions of the bank's return through equity than debt, the probability of bankruptcy, and hence its expected cost, will be reduced.

The collateral value of the firm is also germane here. The more tangible is the form of the firm's investment, the less opportunity is presumably available to the firm to engage in asset substitution that increases risk. The fraction of a firm's assets accounted for by tangible assets is therefore a (negative) indicator of discretionary opportunities to shift risk to lenders. The higher the firm's collateral value, therefore, the lower will be lenders' shareholding. Empirically, however, it will be difficult to untangle this risk-attenuating aspect of collateral from lowered bankruptcy cost considerations discussed earlier.

Application to Japan

Postwar Japan provides a fertile ground to test the hypotheses outlined above. As noted earlier, its legal system has allowed banks to combine corporate lending with equity participation. Moreover, the high reliance of the corporate sector on external funds during the 1960s through the early 1970s suggests that the magnitude of agency costs, and hence the incentive to mitigate them, would have been significant. Testing the hypotheses, however, requires some institutional background on the postwar Japanese financial system and industrial organization.

Japan's financial markets were tightly regulated until the mid-1970s, when gradual deregulation was begun. One of the main objectives of the authorities was to make industrial financing the virtually exclusive preserve of Japan's financial institutions and to limit their number by strictly controlling entry.[15] As a result, Japan's corporate bond market has remained thin and the number of major corporate lenders limited. Excluding government financial institutions, major conduits of funds consist of a dozen city banks, three long-term credit banks, seven trust banks, and large life insurance companies.[16] According to Hodder and Tschoegl (1985) fewer than 30 financial institutions may control over 90 percent of private lending to large industrial firms.

As is the case elsewhere, large Japanese firms typically procure financing from a consortium of lenders. What distinguishes the Japanese loan consortium, however, is the special role played by the lead or "main" bank. Although a precise and steadfast definition cannot be assigned to the term, a typical main bank would be a city bank, which is the largest lender as well as a significant shareholder of a given firm. One important function of the main bank, as Sheard (1989) put it, is to act as a delegated monitor among lenders. It screens and monitors corporate borrowers on behalf of all lenders in a given consortium.[17] The main bank also shoulders most of the burden of reorganization, bail-outs or outright liquidation when a corporate client encounters financial distress.[18]

The main banking system typically occurs within the organizational context of the *keiretsu*, loosely translated as a corporate group. Firms belonging to these groups tend to maintain long term relationships with one another. Intra-*keiretsu* ties are maintained through informal or implicit commitments; they are also manifest in explicit financial commitments of reciprocal shareholding. At the fulcrum of each of these groups are the major city banks. Thus, for example, the Mitsui Bank will most likely serve as the main bank for most firms belonging to the Mitsui *keiretsu*. Major city banks are themselves flanked by two or three closely affiliated financial institutions. For example, the Mitsui group includes at its financial core Mitsui Trust and Banking, Taisho Marine and Fire Insurance, and Mitsui Life Insurance Company. In view of the close coordination that is said to prevail among the *keiretsu* financial institutions, it seems reasonable to regard them collectively as a single economic agent; that is, for purposes of mitigating the agency costs of external finance, it is the *collective* shareholding of all *keiretsu* financial institutions that is likely to matter. I therefore use this level of aggregation to measure the bank's shareholding in the empirical analysis to follow. For ease of exposition, the term "main bank" throughout the remainder of the paper will refer to the **group** of *keiretsu* financial institutions centered around the city bank.[19]

II. An Empirical Test

Variables and Empirical Proxies

Main bank's shareholding. This is the dependent variable in the estimated regression equations. Firm-level data for this variable were compiled from Economic Research Association's (Keizai Chosa Kyogikai) annual publication, *Keiretsu no Kenkyu* (Research in Corporate Financial Groups). ERA employs a primarily quantitative criterion to define a financial group. If a firm has obtained the largest amount of financing from the same bank for three or more consecutive years to date, then that firm is classified as belonging to the bank's *keiretsu*. As in any exercise in taxonomy, ambiguities inevitably arise, and the ERA applies two additional criteria for inclusion where necessary: (i) shareholding by group members exceed 20 percent; (ii) historical ties.

Bankruptcy costs. Two set of proxies were considered. The first is the collateral value of the firm—defined as the proportion of tangible assets in the firm's total assets—as a negative correlate of bankruptcy cost.[20,21] According to Myers (1977), the deadweight losses of bankruptcy will be more pronounced for intangible assets that are linked to the health of the firm as a going concern; that is, the lower the collateral value of the firm, the higher will be its expected bankruptcy costs. For example, technical know-how, human capital, and brand image are likely to lose a greater proportion of their values than tangible physical assets such as plant and equipment, when the firm ceases, or is threatened to cease, as a going concern. If this argument is correct, one should observe a negative relationship between the level of shareholding by lenders and the ratio of the firm's tangible assets to total assets.

As is often the case, however, this empirical proxy may not *uniquely* capture the theoretical attribute we wish to measure. As already noted, collateral value may also serve as a negative correlate of risk. An additional complication concerns the so-called asset specificity effect suggested by Williamson (1985). The idea is that as a firm's assets become more "specific" to existing contractual relations, and hence less redeployable outside these relationships, the salvage value of the firm, *given* that bankruptcy occurs, will decline; that is, bankruptcy costs rise with the degree of asset specificity of the firm and the hypothesized effect is a greater reliance on equity structuring the bank's return through equity participation. One cannot rule out ex ante the possibility that a firm's ratio of fixed to total assets also proxies for the degree of asset specificity. The two will exert opposite effects.

The second proxy tries to capture the indirect cost of bankruptcy. The conflict of interest between shareholders and creditors during financial distress may seriously impair the firm's capacity to take appropriate actions to stem further deterioration. Appropriate action may entail timely disposal of the firm's rapidly depreciating assets or, alternatively, new investment to boost its competitiveness. In either case, the indirect costs of financial distress may be more acute when the firm operates in a very dynamic and rapidly growing market. For one, the failure to keep up with market growth, let alone shutdowns, would exact a high toll in terms of forgone output and, perhaps more importantly, loss in market share. The expected *industry*-wide growth level was therefore included in the regression analyses as a possible correlate of the (expected) indirect cost of financial distress.

External Financing Ratio. This variable is intended to measure the extent to which a firm relies on the main bank to finance its investments. I therefore computed the ratio of financing obtained from the largest *keiretsu* financial group to total assets.

Expected Profitability. The level of profitability was proxied by the rate of business profit defined as:

$$RBP = \frac{\text{gross profit} + \text{receipts of interest plus dividends}}{\text{total assets}},$$

where gross profit is earnings before taxes and interest payments. The choice of this proxy over other measures of profitability is due to Nakatani (1984). He argues that since total assets equal own capital plus debt, the rate of return on total assets should include both current profits and interest paid to financial institutions; counting current profit alone will bias downward the rate of return for those firms with a greater debt burden.

Risk. A natural proxy for risk is a measure of volatility defined as the standard deviation of the rate of business profit.[22] In addition, three other possible proxies of risk were considered. The first is the expected growth of the firm. It has been argued that insiders may have greater potential to obtain information about the prospects of a more rapidly growing firm than outsiders do. A rapidly growing firms may also be riskier because insiders potentially have greater scope and discretion to engage in risky activities; for example, they have greater flexibility in the choice of future investments.[23] The estimated model therefore included expected sales growth as a positive indicator of risk.

Second, the age of the firm was included to explore the possibility that the agency costs are less severe in older firms than in new firms. Compared to relatively mature firms charting familiar waters, newer firms may face greater uncertainty, for example, exploring new technology or markets.[24] Furthermore, established firms may be more cautious about jeopardizing their reputations for the sake of short-term gain through morally hazardous behavior.

Finally, the size of firms may also proxy for risk. Larger firms tend to be more diversified and hence less prone to bankruptcy risk. This implies the opposite effect on the firm's claim structure: banks should hold lower residual claims in larger firms.[25]

Table 1 takes stock of the arguments presented thus far.

Table 1

Predicted effects on the level of shareholding by the main bank

Theoretical variables	Proxies	Predicted effects
Bankruptcy cost	Collateral value	− or +
	Expected industry growth rate	+
External financing ratio	Ratio of Main bank financing to Total asset	+
Expected profitability	Rate of business profit (RBP)	−
Risk	Collateral value	−
	Standard deviation of RBP	+
	Expected growth of the firm	+
	Age of firm	−
	Size of firm	−

Data and Estimation Procedure

The variables were analyzed for the period 1964-1970, because it represents the last major investment boom (56 months of uninterrupted growth from November 1965 to July 1970) before the first oil shock. Data on the sources of loans and shareholder composition are from *Keiretsu no Kenkyu*. The year of the firm's establishment was taken from *Kaisha Nenkan* [Company Annual] published by Nikkei. All other variables were compiled from the NEEDS corporate financial data tape which includes balance sheet, income statement, and other supplementary accounting data at the firm level. Since the financial settlement of the majority of firms was on a semi-annual basis until recently, the flow data were added up component by component. No problem was encountered on the comparability of data across these different sources since all of them were ultimately compiled from the same source, the *Yuka Shoken Hokokusho*—financial reports that all listed companies are required to submit on an annual or semiannual basis to the Ministry of Finance.

The sample is drawn from 635 firms that were continuously listed on the First Section of the Tokyo Stock Exchange (TSE) from 1964 to 1970 inclusive. To exclude regulatory and other related effects, sample selection was limited to firms in manufacturing industries which reduced the sample to 468.[26] The list was further pared down by eliminating firms undergoing mergers or severance during this period, firms with incomplete records on the variables included in the analysis, and firms eliminated from the data tape by Nikkei because they were either involved in mergers or severance in the period following 1970, or because they simply went bankrupt. The sample therefore contains a nonnegligible degree of self-selection with a likely bias toward relatively larger, successful firms.[27] These rounds of elimination left a final tally of 338 firms. Their breakdown according to Nikkei's industry classification is reported in Table 2.

Table 2
Industry classification of sample firms

Industry classification	No. of companies
Food	30
Textiles	29
Pulp and paper products	16
Chemicals	51
Pharmaceutical	14
Oil and coal products	5
Rubber products	4
Glass and ceramics	17
Iron and steel	29
Nonferrous metals	24
Machinery	36
Electric machinery	33
Shipbuilding	5
Automobile	19
Transport equipments	6
Precision instruments	11
Others	9
TOTAL	338

The sampling period was divided into two subperiods—1964-66 and 1967-70—over which sample averages were calculated. Averaging was performed to reduce measurement error due to random year-to-year fluctuations in the variables. The procedure should also serve to smooth the effects of lumpy investments undertaken in a particular year on accounting data. The 3-year averages of the dependent variable, firm size, asset structure, and external financing ratio were measured for the contemporaneous period 1964-67. The variables pertaining to expectations—profitability and growth—were measured over the period 1967 through 1970; that is, I assume rational expectations and use the ex post realized values as proxies of the values expected when the financing decision was made.[28] Finally, the standard deviation of the change in the rate of business profit was measured using all seven years in the sample in order to obtain the maximum efficiency in the measure as possible.

Table 3 reports the summary statistics of the variables analyzed. On average, the level of borrowing from the group of *keiretsu* financial institutions represented a little under 24 percent of the firms' total asset or over 35 percent of total borrowing. A noteworthy finding is that for the vast majority of the sample firms, the legal ceiling of 10 percent did not appear binding: The average main bank's shareholding stood at under 11 percent. (Recall that is a collective measure which typically includes holdings of three or more financial institutions.)

Though not reported in the Table, the second noteworthy finding is that for close to 93 percent of the firms in the sample, the main bank was simultaneously a lender and a shareholder in the firm. Out of a sample of 334 firms, the instance where the main banks held outstanding loans but held no stock was limited to 25 firms (7.5 percent). The prevalent mode of financing for the sample firms thus featured principal lenders as shareholders. Though impressionistic, the finding is consistent with the postulated efficiency of conjoining debt and equity claims.

Our next step is to ascertain whether the cross section data reveal a systematic pattern in the level of shareholding. Since for a small but not negligible proportion of the observations on the dependent variable assumed a limiting

Table 3
Descriptive statistics for the sample firms

Variables	N	Mean	Standard Deviation
Main bank's shareholding	334	0.106	0.125
Asset composition	338	0.247	0.096
External financing ratio	337	0.237	0.112
Ratio of main bank borrowing to total borrowing	338	0.353	0.150
Expected profitability	337	0.207	0.142
Volatility	336	0.035	0.025
Expected growth	337	0.203	0.134
Age (years)	338	34.813	16.121
Size (million Yen)	338	32,274	46,790

Note: all variables except Expected Growth, Age, and Size are ratios; i.e., multiplying them by 100 would convert them to percent levels.

value—i.e., the bank's equity holding was zero—a maximum likelihood Tobit estimator was used.[29] For purposes of comparison, an ordinary least squares estimation was also performed.

The Results

As a preliminary step to running the regressions, a Pearson correlation analysis was performed. The magnitude of collinearity among the explanatory variables was minimal, so the results are not reported. To check for possible industry effects on the model, a standard analysis of variance was performed using industry dummy variables.[30] Surprisingly, no statistically significant industry effect was detected in the data.

The main empirical findings of this paper are contained in Table 4.[31] No material difference is discernible between the estimated coefficients of the TOBIT and OLS regressions. This suggests that the censoring problem was minimal. Finally, different functional specifications were also tried for possible non linear relationships between the variables. The results turned out to be uniformly inferior to the linear specification and are not reported.

To begin with the proxies for bankruptcy cost, the collateral value of the firm yielded a positive coefficient. The model thus predicts that as the proportion of tangible assets in total assets increases, the lender's shareholding in the firm will increase. This clearly runs counter to the commonly subscribed view that tangible assets lose less value relative to intangible assets in times of financial distress. Several interpretations are possible. For example, the obtained sign may reflect the asset-specificity problem discussed by Williamson (1985), i.e., tangible assets, as positive correlates of transaction-specific investment, impose higher bankruptcy costs. However, it is difficult to reconcile this argument and the prediction that higher collateral also means lower risk and hence leads to lower shareholding by the main bank. In light of this, one cannot dismiss the possibility that the "perverse" sign may be due to the measurement problems discussed earlier, particularly with respect to the valuation of the firm's land holdings.

The proxy for indirect bankruptcy costs—the expected growth of the industry— yielded the predicted positive sign; that is, the main bank tends to hold higher equity stakes in firms expecting more rapid growth. The statistical significance of the estimate is rather tentative, however. No doubt, this reflects the considerable amount of statistical noise that is likely to intrude on industry classifications of the sample firms.

The most notable result is the positive and statistically significant coefficient (at the 1 percent level) for the firm's external financing ratio; that is, as larger fractions of the firm's investment are financed by the main bank's funds, the larger is the latter's equity stake in the firm. According to the TOBIT estimate, an increase by 0.111 in the external financing ratio—one standard deviation in the sample—increases the main bank's shareholding in the firm by 2.7 percent.[32] Similar conclusions can be drawn on the basis of OLS results. These findings lend support to an oft-noted observation that Japanese banks exercise considerable influence on the firm by virtue of their shareholding.[33]

Table 4
Determinants of the level of main bank's shareholding

	TOBIT regression	OLS regression
Collateral value	0.197	0.174
	(6.175)**	(2.268)**
Expected industry growth	0.401	0.405
	(3.461)*	(1.962)*
External financing ratio	0.244	0.241
	(14.826)***	(3.934)***
Expected profitability	−0.077	−0.070
	(1.570)	(−1.183)
Volatility	0.665	0.665
	(3.845)**	(2.012)**
Expected firm growth	0.018	−0.011
	(0.442)	(−0.132)
Age	-1.8×10^{-5}	-6.9×10^{-5}
	(0.002)	(−0.164)
Size	-3.1×10^{-7}	-2.5×10^{-7}
	(4.04)**	(−1.73)*
Constant	−0.096	−0.080
	(2.291)	(−1.305)
	Limit observations: 25	
	Non-limit observations: 311	
	Log likelihood: 182.45	
	Squared Correlation (Y,E(Y)): 0.111	R^2: 0.093
	LR test: 38.91***	F-statistics: 5.234***

Note: Significance levels: * = 0.10, ** = 0.05, *** = 0.01; chi-square statistics (t-statistics) in parentheses for TOBIT (OLS) regression.

Although an inverse relation obtained as was predicted between the level of main bank's shareholding and the expected profitability of the firm, the coefficient turned out to be statistically insignificant.

As predicted, a positive relation obtained between the main bank's shareholding and the volatility of the firm's profit which proxied for risk. The estimated relationship predicts that an increase in volatility by the magnitude of one standard deviation in the sample will increase the bank's shareholding in the firm by nearly 1.7 percent. By contrast, the second proxy for risk, the expected growth of the firm did not yield a significant result. Although an inverse relation obtained between the level of the main bank's shareholding and the firm's age, the coefficient was statistically insignificant.[34] However, an inverse and statistically significant relationship did obtain between the firm's size and the percentage of residual claims issued to the bank.

III. Conclusions

On the basis of published accounting data, this paper investigated the benefits of joint debt-equity financing, by examining the determinants of ownership structure of Japanese firms during the era of rapid growth. Any test of optimal financial structure, motivated in part or in whole by information asymmetry, must necessarily be crude. By definition, the researcher who must depend on publicly available information is subject to the very type of information asymmetry faced by a firm's outside investor. In this respect, the student of Japanese corporate finance is especially handicapped. The statistical noise due to the poor quality of corporate disclosure in Japan no doubt accounts for a significant portion of the sizable unexplained variance that remains in the regressions.[35]

These limitations notwithstanding, the empirical analysis did yield plausible results in support of the theoretical model. For one, rarely did banks forgo the opportunity to hold an equity stake in firms where significant lending occured. Thus, if any efficiency gains were made possible by a legal system that conferred securities powers to banks, they did not go unexploited. Furthermore, the cross section evidence generally supports the view that agency cost considerations indeed seem to matter in how banks structure their claims in client firms. The most notable results in this respect are that the amount of equity stake that the bank held in the firm increased with the level of financing the bank extended to the firm and with the level of risk. The bank's capacity to participate in corporate equity is crucial in "reconciling" the aggressive lending behavior of Japanese banks during the period of rapid growth on the one hand, and on the other, the low level of net worth and hence relatively high loan risks in the corporate sector. These considerations may be relevant to policymakers in the U.S. currently faced with the task of overhauling its banking system.

ENDNOTES

1. In the process of dissolving the *zaibatsu*—the financial "cliques" which allegedly precipitated Japan's entry into war—the Occupation forces severely curtailed the power of Japanese banks in addition to outlawing intercorporate shareholding by nonfinancial corporations. These stiff provisions were subsequently relaxed as Japan was poised to embark from reconstruction to rapid peacetime growth. Reciprocal shareholding became legal again in 1949, thus ushering the way for many of the former *zaibatsu* firms to regroup under the present day *keiretsu*. And unlike in the United States, the city banks which form the nuclei of these corporate groups, (along with trust banks and insurance companies) were empowered to hold corporate shares subject to a legal maximum. Until 1987, the legal limit was set to 10 percent of the outstanding stocks of any single company; the ceiling currently stands at 5 percent.

2. This very question was addressed also by Pozdena (1991).

3. See for example, Kester (1986), Allen and Mizuno (1989) and Prowse (1990).

4. This section draws heavily from Kim (1991) which provides a more formal treatment on the subject. A compressed version of the formal model is sketched out in Appendix A.

5. One can think of this variable as unforeseen events such as bad weather, a technological discovery, or war in the Persian Gulf.

6. This assumption abstracts the bond market (as well non-financial credit intermediaries) from the analysis, thus allowing us to focus on the implication of granting equity powers to financial intermediaries. The implicit premise here is that in the economy under consideration, intermediated debt finance (i.e., bank loans) are preferred to direct borrowing (bonds). This could be because, as is now widely recognized in the literature, banks possess a comparative advantage in information production, such as screening and monitoring borrowers. By implication, this comparative advantage will be heightened in situations where information-related problems are acute in the system, as is presumed here.

7. Empirical analysis on bankruptcy costs is scarce. Frequently cited is a case study of a railroad bankruptcy by Warner (1977) which finds the direct costs to be rather negligible. But a more recent case study of corporate reorganization in the oil industry by Cutler and Summers (1988) suggests that the indirect costs of financial distress can be very substantial indeed. According to the authors' estimate, the dispute between Texaco and Pennzoil over the Getty Oil takeover imposed a deadweight loss equivalent to some one-sixth of the combined wealth of the two companies.

8. We say residual profit because the profit accruing to shareholders consists of what remains after creditors have been paid off. For this reason, equity is sometimes referred to as a "residual claim" to the firm. By contrast, debt is a "fixed claim" since its return, provided the firm is solvent, is invariant to the firm's profit; if the firm is insolvent, fixed claimants become residual claimants, i.e., creditors take over the ownership of the firm's assets.

9. It is for these reasons that the distinction between the original and new shareholders is crucial. To reflect their different access to information, the initial owner will sometimes be referred to as the "inside share" and the new shareholder as "outside share."

10. The distinction between the inside and outside shareholder is not important here. The outside shareholder will end up participating in the moral hazard of risk shifting to debtholders, without necessarily being privy to the insider's information or his action.

11. The modern theory of financial intermediation critically hinges on the postulated efficiency of banks as monitors in the presence of moral hazard. For example, Diamond (1984) argues that diversification within the intermediary enables it to monitor at a lower cost than if scattered principals (depositors) were to monitor individually. In a related vein, in view of the bank's superior access to the firm's information, Fama (1985) distinguishes bank loans as "inside debt" from the open capital market "outside debt" such as bonds.

12. I focus on the level of equity holding by Japanese banks and omit discussion on the level of debt for the following reason. Under the rubric of the so-called "low interest rate policy," a usury regulation on corporate lending remained in effect throughout most of the postwar period. There is now ample evidence that banks tried to circumvent this regulation, most notably by requiring firms to post compensating balances. Reported interest rates are likely to diverge from the effective interest rate and hence, measured levels of debt may diverge substantially from actual levels. (See, e.g., Wakita (1983)).

13. By implication, a firm will try to finance an investment project out of internal funds whenever possible and thereby avoid the deadweight costs of going to the external market. This option will be foreclosed, however, if investment projects are "lumpy," i.e., project size rises in discrete increments and the minimum.

14. Section II provides a discussion on the empirical proxies for this and other firm-specific parameters.

15. For a concise survey of Japanese financial markets and corporate finance, see Hodder and Tschoegl (1985).

16. The concentration ratio remains high in this industry. At the end of 1980s, the top eight domestic life insurance companies among a total of 21 controlled over 80 percent of total assets (Hodder and Tschoegl 1985, p. 176).

17. See also Horiuchi et al., (1988), Horiuchi (1989) and Hoshi et al., (1990a).

18. See Hodder and Tschoegl (1985), Suzuki and Wright (1985) and Hoshi et.al. (1990b). According to Nakatani

(1984), the main raison d'être of *keiretsu* and main banking is precisely to minimize the probability of encountering financial distress through a mutual risk-sharing arrangement among member firms and financial institutions.

19. The membership of the various *keiretsu* financial institutions as well as a list of "independent" ones, i.e., financial institutions without any *keiretsu* affiliation, are in Appendix B.

20. Another proxy for bankruptcy costs (as well as risk) often used in the literature is the level of R&D and advertising expenditures as a proportion of the firm's total sales or total assets. The presumption here is that these variables measure the firm's growth opportunities, i.e. they add value to the firm but cannot be collateralized. Unfortunately, Japanese companies were not required to disclose expenditures on these items until the 1970s; hence this particular proxy could not be included in the estimated model.

21. Total tangible fixed assets consists of depreciable and nondepreciable assets. Buildings and structures, machinery and equipment, vessels and vehicles, etc., fall under the former category; land and construction in progress fall under the latter. Inflation accounting was virtually unheard of in Japan during the period under review. Hence, assets reported in the balance sheet understate their prevailing value by a significant margin. Corporate assets held in land are particularly problematic given the steep increase in real estate prices during the postwar period. This is doubly troublesome since land has been one of the traditionally favored forms of collateral required by banks. In the absence of more detailed information on corporate land holding, any attempt at market value adjustment, however, is likely to introduce additional measurement error. The empirical model was therefore estimated using book values of assets.

22. This is a scaled measure of volatility, since business profit was normalized by the firm's total assets.

23. See for example MacKie-Mason (1989) and Titman and Wessels (1988).

24. These are some of the very reasons why *de novo* firms often obtain financing through joint ventures rather than bank loans.

25. The size of a firm will also have direct bearing on how much bargaining power it has vis-à-vis the bank. If Japan's banking industries was not perfectly competitive, greater bargaining power for the firm may imply lower levels of residual and fixed claims issued to the bank.

26. Nonmanufacturing industries, such as communications or utilities, tend to be heavily regulated. Corporate financing decisions will not be neutral to regulations. For example, holding other things constant, a firm operating in an industry which limits entry will be able to support a greater amount of debt because entry barriers confer oligopoly rent and lowers business risk. Regulated industries are also subject to greater public scrutiny and hence are more limited in engaging in moral hazard.

27. Admittedly, the moral hazard problem for these larger firm would be much less severe than for smaller firms. Unfortunately, as is the case elsewhere, access to data about smaller (and hence unlisted) Japanese firms is very limited. Criticism of self-selection bias should be tempered on at least one count, however, namely, that the dependence on external financing even for the largest firms was very pronounced during the period under review. On this ground alone, one cannot dismiss the potential moral hazard problem as trivial. The proof of the pudding of course is in the eating.

28. This also obviates possible multicollinearity problems between profitability and external financing requirement. *Ceteris paribus*, external financing ratio are lower for more profitable firms since larger fractions of investments can be financed through retained earnings. The collinearity is avoided when forward values of profitability are used.

29. The TOBIT technique is designed to use all observations, both those at the limit (in our case, bank's shareholding equaling zero) and those above it, to estimate a regression line. Tobin pioneered this technique in his classic study of the influence income on household expenditures on durable goods, where a large percentage of the sampled households made no durable purchases during the survey period. The idea is that since durable goods by nature are not divisible, a certain threshold level of income must to be reached before one actually observes positive levels of purchase. The coefficient estimated by TOBIT thus explains the change in the dependent variable y in terms of two components: (1) the change in y of those above zero, weighted by the probability of y being greater than zero; and (2) the change in the probability of y being greater than zero, weighted by the expected value of y given that it is greater than zero. Generally speaking, when the dependent variable is censored, OLS estimation will yield biased estimates of coefficients. How significant this bias is will depend on the severity of the censoring problem.

30. The data problem discussed in footnote 20 provides one motivation. One would expect to find systematic interindustry differences in the relative importance of R&D and advertising expenditures: for example, pharmaceutical firms are likely to be more research-intensive than textiles manufacturers. Any systematic variation that remains in the error term due to the omission of this variable will be picked up by the industry dummy.

31. Industry dummies were not included since they were found to be insignificant.

32. Strictly speaking, the coefficient from TOBIT estimation overstates the effect of the explanatory variables. The appropriate procedure is to weigh the coefficient by the expected probability that a given observation has a nonlimiting dependent variable. Since this weight turned out to be virtually equal to unity, little harm is done by ignoring this caveat. For a more extensive discussion on this issue, see McDonald and Moffitt (1980).

33. For example, Japanese banks are said to wield much influence on client firms' capital spending plans. Particularly noteworthy in regard to banks' corporate control through shareholding is the legal provision that allows major shareholders to remove corporate directors at any time, without cause by an ordinary resolution of a shareholder general meeting.

34. Regressions were also run with the age variable specified as a dummy variable, taking a value of 1 if the firm was founded after W.W.II and 0 otherwise. This did not yield a significant estimate either.

35. The poor quality of accounting data in Japan is not an aberration; it should be expected. Throughout most of the postwar period, corporate finance was the virtually exclusive domain of banks. By implication, banks monopolized on corporate monitoring and used the information in allocating investable funds. In the absence of any active open issue market to speak of, the lack of public disclosure as deep and broad as found in the U.S. or U.K. is natural to expect.

36. To minimize notational clutter, the firm-specific parameters μ and α will be suppressed unless called for in the analysis.

37. Note that this amounts to saying that equity issues do not in themselves generate funds. This simplifying assumption is actually consistent with Japanese corporate financing practice during the period under review. Virtually without exception, equity was issued at a par value of 50 yen per share. Needless to say, this issue price represented a negligible fraction of the market value of equity for most of the listed firms.

Appendix A
A Sketch of a Model of Optimal Financial Contracts

This Appendix outlines a simple model of optimal financial contracts between an entrepreneurial firm and a shareholder bank. The return to the firm's investment takes the form $\tilde{X} = e + \theta + \mu$; e denotes effort, θ is a random variable, and μ is a firm-specific parameter that indexes the profitability of investment. Return increases in all three variables. Unless stated otherwise, assume that θ is uniformly distributed over $[\underline{\theta}, \bar{\theta}]$ with $E\theta = 0$. Denote the spread of the distribution (hence the riskiness of the investment) by α, $\alpha = \bar{\theta} - \underline{\theta} = 2\bar{\theta}$.[36]

To procure some fixed amount of external financing, L, the owner-manager can issue the bank a fixed claim (i.e. debt) equal to RL, where L is the amount borrowed and $R = (1 + r)$ is the gross interest rate, and/or cede a portion of his equity λ, retaining for himself the remaining fraction $(1 - \lambda)$.[37]

Assuming risk-neutrality throughout, the expected profit of the owner-manager is:

$$\text{(A1)} \quad \phi = (1-\lambda)\int_{\theta^*}^{\bar{\theta}} [e + \theta - RL]dF(\theta) - u(e),$$

where $\theta^* = RL - e$ denotes the critical value of θ below which bankruptcy occurs, and $u(e)$ measures the disutility that the manager associates with effort. Assume $u_e > 0$, $u_{ee} > 0$. Note that because of limited liability, the owner-manager cares about the expected return only over the states where the firm is solvent.

The expected profit of the bank is given by:

$$\text{(A2)} \quad \pi = RL(1-F(\theta^*)) + \lambda\int_{\theta^*}^{\bar{\theta}} [e+\theta-RL]dF(\theta) + \int_{\underline{\theta}}^{\theta^*} [e+\theta-B]dF(\theta) - \rho L \geqslant 0.$$

The first term is the debt owed to the bank times the probability that the firm will be able to honor it. The second term is the return accruing to the bank through its equity holding, while the third is the expected value of the firm over states where the firm defaults. Bankruptcy is costly in the sense that the value of the firm is eroded by some fixed amount B when the bank takes over ownership. Finally, the last term represents the opportunity cost the bank associates with providing L to the firm. Competition among banks ensures that equation (2) equals zero.

For any given combination (R, λ), the manager will set the effort level so as to maximize his own expected profit; that is, e^* will be chosen to satisfy:

$$\text{(A3)} \quad \phi_e = (1-\lambda)(1-F(\theta^*)) - u_e = 0.$$

To ensure a maximum, assume $\phi_{ee} < 0$. Implicit differentiation of (A3) yields the direction of adjustment in effort with respect to the contract variables:

$$\text{(A4)} \quad \begin{aligned} e^*_R &= (1-\lambda)Lf(\theta^*)/\phi_{ee} < 0 \\ e^*_\lambda &= (1-F(\theta^*))/\phi_{ee} < 0 \end{aligned}$$

In valuating the claims issued by the firm, the rational bank will anticipate these effort responses in addition to changes in the expected cost of bankruptcy. Consequently, the problem facing the owner-manager is to choose a structure of claims R (and hence RL since L is fixed) and λ that maximizes his expected profit (1) subject to the "budget constraint" (A2) and the incentive constraints (A3) and (A4). The relevant first-order condition for an interior optimum is given by:

$$\text{(A5)} \quad \frac{\phi_\lambda}{\phi_R} = \frac{\pi_e e_\lambda}{\pi_e e_R - BLf}; \; \pi_e = \lambda(1-F)+F+Bf > 0.$$

The left-hand side (LHS) of (A5) is the ratio of the marginal response in the firm's profit with respect to the contract variables, $\phi_R < 0$, $\phi_\lambda < 0$. The RHS is the ratio of the marginal deadweight costs. For issues of equity claims (λ), this deadweight cost equals the marginal response of the bank's expected profit with respect to effort times the change (decline) in effort induced by an increase in λ. For issues of debt (R), the deadweight cost divides into two components: $\pi_e e_R < 0$, which has a similar interpretation as above, and BLf, the marginal increase in the expected cost of bankruptcy.

Because the optimization problem involves a nonlinear constraint, the comparative static analysis turns out to be quite complex. The discussion here highlights the basic intuition underlying the adjustments in the equilibrium value of the bank's shareholding λ^*, in response to changes in firm-specific parameters. (See Kim (1991) for a more complete treatment). From (A5), λ^* will depend on the severity of the deadweight cost of equity *relative* to that of debt. All other things equal, therefore, the direction of adjustment in λ^* will depend on how a given parameter affects the relative severity of these (marginal) deadweight costs.

Consider first an increase in bankruptcy cost B which is the most transparent case. Intuition suggests that this increases the deadweight cost of debt relative to that of equity and hence leads to the adjustment $\partial\lambda^*/\partial B > 0$. A necessary condition for this to obtain is:

$$\text{(A6)}\quad \frac{\partial}{\partial B}\left(\frac{\pi_e\, e_\lambda}{\pi_e\, e_R - BLf}\right) < 0 \rightarrow (\pi_e\, e_R - BLf)(e_\lambda f) - \pi_e\, e_\lambda (e_R f - Lf) < 0.$$

Rearranging and simplifying yields the inequality $L[\lambda(1 - F(\theta^*)) + F + Bf] > BLf$, and hence $\partial\lambda^*/\partial B > 0$.

Following a similar procedure, the direction of adjustment for λ^* with respect to an increase in L, the extent to which the firm relies on the bank's funds, will hinge on the sign of

$$\text{(A7)}\quad (\pi_e\, e_R - BLf)[\pi_e\, e_{\lambda L} + e_\lambda(\pi_{eL|\bar{e}} + \pi_{ee}\, e_L)] - \pi_e\, e_\lambda[\pi_e\, e_{RL} + e_R(\pi_{eL|\bar{e}} + \pi_{ee}\, e_L) - Bf].$$

With some tedious algebra it can be shown that e_{RL}, e_L, $\pi_{ee} < 0$, $\pi_{eL|\bar{e}} > 0$ (hence $\pi_e\, e_\lambda[\bullet] > 0$), and that $\pi_e\, e_{\lambda L} + e_\lambda(\pi_{eL|\bar{e}} + \pi_{ee}\, e_L) \geqslant 0$ for all $\pi_e \geqslant u_e$ (which holds for all cases of interest). Therefore, the expression in (A7) is negative, which in turn implies $\partial\lambda^*/\partial L > 0$.

In contrast to the previous two cases, an increase in the profitability of investment μ increases the relative deadweight cost of equity and hence leads to a lower λ^*. A necessary condition for such an adjustment to obtain is:

$$\text{(A8)}\quad (\pi_e\, e_R - BLf)[\pi_e\, e_{\lambda\mu} + e_\lambda(\pi_{e\mu|\bar{e}} + \pi_{ee}\, e_\mu)] - \pi_e\, e_\lambda[\pi_e\, e_{R\mu} + e_R(\pi_{e\mu|\bar{e}} + \pi_{ee}\, e_\mu)] > 0.$$

It is relatively straightforward to show that $e_\mu > 0$, $e_{R\mu} = 0$, $e_{\lambda\mu}$, $\pi_{e\mu|\bar{e}} < 0$ and hence that $\pi_e\, e_\lambda[\bullet] < 0$. A negative (or zero) value for the expression inside the first set of square brackets is thus sufficient to establish the inequality in (A8). After some manipulation this can be shown to hold unambiguously for cases where $\pi_{ee} > \phi_{ee}$, i.e., the marginal return to effort diminishes faster for the firm than the bank.

Finally, intuition suggests that increased riskiness should lead to a greater issue of λ^*. It turns out, however, that the actual direction of adjustment is sensitive to the distributional assumption as well as to the initial equilibrium value of λ^* and R^* (and hence to the initial default probability). Full treatment on this issue exceeds the scope of this paper. Instead, we simply state the result here that $\partial\lambda^*/\partial\alpha > 0$ obtains with the least degree of ambiguity for cases $\partial f(\theta^*)/\partial\alpha > 0$ where α indexes the riskiness of the distribution. The intuition is simple. All other things equal, the relative severity of deadweight cost associated with debt rises if greater riskiness increases the marginal deadweight cost of bankruptcy, BLf.

Appendix B
Keiretsu Affiliations of Financial Institutions, 1964–70

Affiliated Banks

Mitsui Group:
Mitsui Bank, Mitsui Trust and Banking, Taisho Marine and Fire Insurance, Mitsui Life Insurance Company.

Mitsubishi Group:
Mitsubishi Bank, Mitsubishi Trust and Banking, Tokio Marine and Fire Insurance, Meiji Life Insurance Company.

Sumitomo Group:
Sumitomo Bank, Sumitomo Trust and Banking, Sumitomo Marine and Fire Insurance, Sumitomo Life Insurance Company.

Fuyo (Fuji) Group:
Fuji Bank, Yasuda Trust and Banking, Yasuda Fire and Marine Insurance, Yasuda Life Insurance Company.

Sanwa Group:
Sanwa Bank, Toyo Trust and Banking, Daido Life Insurance Company.

Dai-Ichi Group:
Dai-Ichi Bank, Asahi Life Insurance Company.

Unaffiliated Banks

Long-Term Credit Banks:
Industrial Bank of Japan, Long-Term Credit Bank of Japan, Nippon Credit Bank

City Banks:
Hokkaido Takushoku Bank, Bank of Tokyo, Daiwa Bank, Tokai Bank, Kyowa Bank, Kobe Bank, Nippon Fudosan Bank, Norin Chuo Kinko

REFERENCES

Allen, D.E., and H. Mizuno. 1989. "The Determinants of Corporate Capital Structure: Japanese Evidence." *Applied Economics* 21, pp. 569-585.

Cutler, D., and L. Summers. 1988. "The Cost of Conflict Resolution and Financial Distress: Evidence from the Texaco-Pennzoil Litigation." *Rand Journal of Economics* 19, pp. 157-172.

Diamond, P. 1984. "Financial Intermediation and Delegated Monitoring." *Review of Economic Studies* 51, pp. 393-414.

Fama, E. 1985. "What's Different About Banks?" *Journal of Monetary Economics* 15, pp. 29-39.

Hodder, J. and A. Tschoegl. 1985. "Some Aspects of Japanese Corporate Finance." *Journal of Financial and Quantitative Analysis* 20 (June) pp. 173-191.

Horiuchi, A., F. Packer, and S. Fukuda. 1988. "What Role Has the 'Main Bank' Played in Japan?" *Journal of the Japanese and International Economies* 2, pp. 159-180.

Horiuchi, A. 1989. "Informational Properties of the Japanese Financial System." *Japan and the World Economy* 1, pp. 255-278.

Hoshi, T., A. Kashyap and D. Scharfstein. 1990a. "Corporate Structure, Liquidity, and Investment: Evidence from Japanese Industrial Groups." *Quarterly Journal of Economics* (February) pp. 33-60.

__________. 1990b. "The Role of Banks in Reducing the Cost of Financial Distress in Japan." *Journal of Financial Economics* 27, pp. 67-88.

Kester, K. 1986. "Capital and Ownership Structure: A Comparison of the United States and Japanese Manufacturing Corporations." *Financial Management* 15 (Spring) pp. 5-16.

Kim, S.B. 1991 "Modus Operandi of Lender-cum-Shareholding Banks." Working Paper. Federal Reserve Bank of San Francisco.

MacKie-Mason, J. 1989. "Do Firms Care Who Provides Their Financing?" NBER Working Paper No. 3039.

McDonald, J.F. and R.A. Moffitt. 1980. "The Uses of Tobit Analysis." *Review of Economics and Statistics* 62, pp. 318-321.

Modigliani, F., and M. Miller. 1958. "The Cost Capital, Corporation Finance and the Theory of Investment." *American Economic Review* 48, pp. 261-97.

Myers, S.C. 1977. "Determinants of Corporate Borrowing." *Journal of Financial Economics* 5, pp. 147-175.

Nakatani, I. 1984. "The Economic Role of Financial Corporate Grouping." In *The Economic Analysis of the Japanese Firm*, ed. M. Aoki. New York: North Holland.

Pozdena, R. 1991. "Why Banks Need Commerce Powers." Federal Reserve Bank of San Francisco *Economic Review* (Summer) pp. 18-31.

Prowse, S. 1990. "Institutional Investment Patterns and Corporate Financial Behavior in the United States and Japan." *Journal of Financial Economics* 27, pp. 43-66.

Sheard, P. 1989. "The Main Bank System and Corporate Monitoring and Control in Japan." *Journal of Economic Behavior and Organization* 11, pp. 399-422.

Suzuki, S. and R. Wright. 1985. "Financial Structure and Bankruptcy Risk in Japanese Companies." *Journal of International Business Studies* pp. 97-110.

Titman, S., and R. Wessels. (1988). "The Determinants of Capital Structure Choice." *Journal of Finance* 43 (March) pp. 1-19.

Wakita, Y. 1983. "Wagakuni no Kashidashi Shijo to Keiyaku Torihiki" [Loan Markets and Contracts in Japan]. *Kin'yu Kenkyu* 2, pp. 47-76.

Warner, J. 1977. "Bankruptcy Costs: Some Evidence" *Journal of Finance* 32 (May) pp. 337-347.

Williamson, O. 1985. *The Economic Institutions of Capitalism.* New York: Free Press.