Problem Loans and Cost Efficiency in Commercial Banks

Allen N. Berger
Board of Governors of the Federal Reserve System

Robert DeYoung
Office of the Comptroller of the Currency

Office of the Comptroller of the Currency
Economic and Policy Analysis Working Paper 95-5
December 1995

The authors thank Bob Avery, Paul Bauer, Gary Ferrier, Kevin Jacques, Julapa Jagtiani, Bill Lang, Tom Lutton, Joe Mason, Andrea Resti, Bernie Shull, and Gary Whalen for very helpful comments; Maria Montano and Joe Scalise for exceptional research support; and Claire Emory for editorial assistance.

Please address questions to Robert DeYoung, Bank Research Division, Office of the Comptroller of the Currency, Washington, DC 20219, phone (202) 874-5427, fax (202) 874-5394, e-mail robert.deyoung@occ.treas.gov.

* Any whole or partial reproduction of material in this paper should include the following citation: Berger, Allen N., and Robert DeYoung, "Problem Loans and Cost Efficiency in Commercial Banks," Office of the Comptroller of the Currency, E&PA Working Paper 95-5, December 1995.

I. Introduction

Over the last several years, two strands of research in the field of financial institutions have received a great amount of attention. One is the issue of problem loans. Virtually all research on the causes of bank and thrift failures find that failing institutions have large proportions of nonperforming loans (past due and nonaccrual) prior to failure. Demirguc-Kunt (1989) reported that the majority of pre-1990s bank failure studies found asset quality to be a statistically significant predictor of insolvency. More recent studies (e.g., Whalen 1991, Barr and Siems 1994) have reinforced this finding.

A second strand of literature deals with the productive efficiency of financial institutions. A number of studies have found that most institutions have high costs and low profits relative to institutions on the "best-practice" efficient frontier. Various studies of mergers, agency problems, corporate governance, branching strategies, foreign ownership, etc. provide evidence that estimates of frontier efficiency are significantly related to the structure, conduct, and performance of financial institutions. Berger, Hunter, and Timme (1993) reviewed much of this literature.

What is largely missing from the literature is an analysis of the intersection of these two strands of research -- that is, the relationships between problem loans and cost efficiency at financial institutions. On the surface, these two topics might appear to be unrelated, because operations personnel typically do not participate in screening and monitoring loan customers, and because loan officers and review personnel typically do not participate in overseeing operations costs. Despite this apparent dichotomy, issues of problem loans and cost efficiency are in fact related in several important ways. The recent empirical literature suggests at least three significant links between these two topics.

First, a number of researchers have found that failing banks tend to be very cost inefficient, that is, located far from the best-practice frontier (see, for example, Berger and Humphrey 1992; Coyne, McManus, and Stagliano 1993; Barr and Siems 1994; Wheelock and Wilson 1994; Becher, DeYoung, and Lutton 1995). Thus, in addition to having high ratios of problem loans, banks approaching failure also tend to have low levels of measured cost efficiency. DeYoung and Whalen (1994) showed that declines in average cost efficiency tend to precede increases in average problem loans for banks approaching failure, while a number of other studies have found negative contemporaneous relationships between efficiency and problem loans even among banks that do not fail (Eisenbeis and Kwan 1994, Hughes and Moon 1995, Resti 1995). Cost-inefficient banks may tend to have loan performance problems for a number of reasons. For example, banks with poor senior management may have problems in monitoring both their costs *and* their loan customers, with the losses of capital generated by both these phenomena potentially leading to failure. We will refer to this as the "bad management" hypothesis. Alternatively, loan quality problems may be caused by an event exogenous to the bank, such as an unanticipated regional economic downturn. The expenses associated with the nonperforming loans that result (e.g., extra monitoring, negotiating workout arrangements, seizing and disposing of collateral, diverted managerial focus) can create the *appearance*, if not the reality, of low cost efficiency. We will refer to this as the 'bad luck' hypothesis. These and other hypotheses linking problem loans to cost efficiency are investigated in the next section.

The second empirical link between problem loans and productive efficiency appears in studies that use supervisory examination data. Peristiani (1993) and DeYoung (forthcoming) both found measured cost efficiency to be positively related to examiners' ratings of bank management quality. Moreover, the latter study found that banks' management quality ratings were more strongly related to their asset quality ratings than to any of their other examination ratings.[1] A relationship between asset quality and cost efficiency (via management quality) is consistent with the failed bank data cited above, and suggests that the negative relationship between problem loans and cost efficiency holds for the population of banks as a whole as well as for failing banks.

Third, some recent studies of bank efficiency have directly included measures of nonperforming loans in cost or production relationships. The stated purposes of this adjustment are to control for the extra costs associated with nonperforming loans and/or to control for underwriting and monitoring expenditures that influence loan quality. Berg, Førsund, and Jansen (1992) made the original observation and included nonperforming loans in a nonparametric study of bank production. Hughes and Mester (1993), Mester (1994a, 1994b), and Hughes, Lang, Mester, and Moon (1995) applied the concept to parametric estimations of bank cost functions and efficiency. As discussed at length in the next section, whether this procedure improves or hinders the estimation of cost efficiency depends upon the underlying reason for the relationship between costs and nonperforming loans.

Thus, a number of important policy and research issues rest on identifying the underlying relationship between problem loans and measured cost efficiency. These include discovering the primary cause of problem loans and bank failures (e.g., external shocks, poor management practices,

---

[1] DeYoung (forthcoming) regressed management quality ratings on examination ratings for capital adequacy, asset quality, earnings quality, and liquidity. Although all of the exam components were significantly and positively related to management quality, asset quality had the largest coefficient value.

or a purposeful lack of managerial oversight); determining the most important supervisory focus for promoting safety and soundness at banks (e.g., loan quality, portfolio diversification, management quality, or cost efficiency); and deciding how to estimate the cost efficiency of financial institutions (with or without controls for problem loans). While we obviously cannot provide definitive answers for all of these important questions, we attempt to shed some light on these issues by analyzing an empirical model that tests a number of competing and complementary hypotheses involving the inter-temporal relationships among problem loans, cost efficiency, and financial capital.

We conduct tests of four hypotheses using Granger-causality analysis. To our knowledge, this paper contains the first tests of any type of the “bad luck,” “bad management,” and “skimping” hypotheses (described in full below), as well as the first Granger-causality test of the well-known “moral hazard” hypothesis. Granger-causality analysis can distinguish between these alternative cases in a way that standard cross-section econometric analysis cannot. This is because several of these hypotheses yield identical predictions for the contemporaneous relationships between problem loans and cost efficiency, but strikingly different predictions for the inter-temporal patterns. Although application of the Granger-causality methodology to microeconomic banking data is unusual, it is not unprecedented (see Berger 1995). We acknowledge that the inter-temporal relationships that we measure are gross statistical associations that do not necessarily prove economic causation. Nevertheless, they can indicate which among the alternative hypotheses is consistent with the data. In the process of our analysis we also find some interesting results concerning the estimation of cost efficiency. Specifically, we find that measured cost efficiency varies substantially with the degree of flexibility in the specification of the cost equation and the error structure.

The empirical results suggest that the inter-temporal statistical relationships between problem loans and cost efficiency run in both directions for U.S. commercial banks between 1985 and 1994. The data suggest that high levels of nonperforming loans Granger-cause or predict reductions in measured cost efficiency, possibly reflecting the extra costs of administering problem loans. This result supports the bad luck hypothesis. Low levels of cost efficiency are also found to Granger-cause or predict increases in nonperforming loans, suggesting that cost-inefficient managers are also poor loan managers. This result supports the bad management hypothesis. The moral hazard and skimping hypotheses are only consistent with subsets of the data.

The remainder of the paper is organized as follows. Section II presents our four hypotheses, identifies their key empirical implications, and discusses their consequences for policy and research. Section III describes the econometric model, and Section IV explains how the efficiency estimates were constructed. Section V describes the data, followed by the empirical results in Section VI. Economic and policy conclusions are addressed in Section VII.

II. Alternative Hypotheses for the Relationship between Problem Loans and Measured Cost Efficiency

Here we outline our four hypotheses -- bad luck, bad management, skimping, and moral hazard -- each of which implies a different inter-temporal relationship in the data. These hypotheses are not mutually exclusive, and any one of the four could dominate the behavior of a given subset of banks. The empirical analysis that follows is designed to determine which among these hypotheses are relatively important for the banking industry as a whole.

Figure 1 displays the expected time patterns of the variables predicted by the hypotheses. For convenience, we will couch all of the descriptions in terms of increases in problem loans, which is the most interesting case for policy purposes. All of the arguments could similarly be made for the less interesting case in which problem loans decrease.

Bad Luck

Under the bad luck hypothesis shown in Appendix A, external events (e.g., a regional recession) precipitate an increase in problem loans for the bank. The important distinguishing implication of the bad luck hypothesis is the relationship between problem loans and measured cost efficiency. After the loans become past due or nonaccruing, the bank begins to expend additional managerial effort and expense dealing with these problem loans. These extra operating costs include, but are not limited to (1) the additional monitoring of the delinquent borrowers and the value of their collateral, (2) the expense of analyzing and negotiating possible workout arrangements, (3) the costs of seizing, maintaining, and eventually disposing of collateral if default later occurs[2], (4) the additional costs of defending the bank's safety and soundness record to bank supervisors during future examinations, and (5) the diversion of managerial attention away from solving other operations problems. Most of these costs, especially the costs associated with loan workout and default, are incurred well after the loan becomes past due. Under the bad luck hypothesis, therefore, significant cost increases are expected to occur sometime after the increase in problem loans. Thus, we expect increases in nonperforming loans to Granger-cause, i.e., temporally precede, decreases in measured cost efficiency under this hypothesis.

It is important to note that under the bad luck hypothesis the extra expenses of dealing with problem loans will create the *appearance*, although not the reality, of lower cost efficiency. Faced

---

[2] An example of this are the expenses that a bank incurs to manage and eventually dispose of other real estate owned (OREO) that was seized as collateral for failed loans.

with an increase in nonperforming loans not of their own doing, even the most cost efficient banks will have to purchase the additional inputs necessary to produce the loan monitoring outputs required to administer these problem credits.

Bad Management

Under the bad management hypothesis shown in Appendix A, poor bank management is the catalyst for a negative relationship between problem loans and measured cost efficiency. Under this hypothesis, low measured cost efficiency is a signal of bad management practices. Subpar managers do not sufficiently monitor their expenses, which results in low cost efficiency, and also do not practice adequate loan underwriting and monitoring. As "bad" managers, they may (1) have poor skills in credit scoring and therefore choose a relatively high proportion of loans with low or negative net present values, (2) be less than fully competent in appraising the value of collateral pledged against the loans, and (3) have difficulty monitoring the borrowers after the loans are issued to assure that covenants are obeyed and the funds are prudentially invested. These underwriting and monitoring problems eventually lead to high numbers of nonperforming loans as borrowers fall behind in their loan repayments.

The key identifying prediction of the bad management hypothesis is the temporal relationship between problem loans and measured cost efficiency, which has the opposite ordering from that predicted by the bad luck hypothesis. Low cost efficiency is expected to precede or Granger-cause higher nonperforming loans, because insufficient cost controls should reduce measured cost efficiency almost immediately, whereas poor loan management will become apparent only after time passes, the loan portfolio becomes seasoned and delinquencies begin to mount. This delay will be exacerbated if managers do not immediately record nonperforming loans on the balance sheet.[3] For example, a bank may lend the borrower additional funds to repay principal and interest on existing loans, thus avoiding having to report problem credits as nonperforming loans.

Skimping

The 'skimping' hypothesis is shown in panel C of Appendix A. This hypothesis is based on an insight by Berg, Førsund, and Jansen (1992) -- which has been developed further by Hughes and Mester (1993), Mester (1994a, 1994b), and Hughes, Lang, Mester, and Moon (1995) -- that the amount of resources allocated to loan underwriting and monitoring may affect both loan quality and measured cost efficiency.[4] Here, the critical decision of the bank lies in the tradeoff between short-term operating costs and future loan performance problems. A bank maximizing long-run profits may rationally choose to have lower costs in the short run by skimping on the resources devoted to underwriting and monitoring loans, but bear the consequences of greater loan performance problems and the possible costs of dealing with these problems in the future. The lower amount of effort devoted to screening loan customers, appraising collateral, and monitoring borrowers after loans are issued makes the bank *appear* to be cost efficient in the short run because fewer operating expenses can support the same quantity of loans and other outputs. The stock of nonperforming loans likely remains unaffected in the short run, because it takes time before borrowers fall behind in their repayments. However, the inattention to the loan portfolio becomes apparent later on after a

---

[3] The General Accounting Office (1990) found evidence that troubled banks systematically underreport nonperforming loans. Anecdotal evidence to this effect has also been reported in the financial press (e.g., *Wall Street Journal*, April 25, 1994).

[4] It is common practice to use nonperforming loan ratios as a proxy for loan quality, chiefly because other dimensions of service and product quality are difficult to measure. One exception is Faulhaber (1995), who uses survey data to show that increases in customer satisfaction are with slightly increased costs.

relatively high proportion of borrowers have become delinquent on their loans.

Under the skimping hypothesis, the Granger-causality between measured cost efficiency and nonperforming loans has the same temporal ordering as the bad management hypothesis. However, it has the opposite sign -- skimping implies a *positive* Granger-causation from measured efficiency to problem loans.

Moral Hazard

The moral hazard hypothesis shown in Appendix A is the classical problem of excessive risk-taking when another party is bearing part of the risk and cannot easily charge for that risk.[5] Under this hypothesis, a loss of earnings from any source -- including the effects of the other three hypotheses shown in Appendix A -- reduces the bank's financial capital. Once capital has eroded, the bank responds to moral hazard incentives by increasing the riskiness of its loan portfolio, which results in higher nonperforming loans on average in the future. Thus, we identify the moral hazard hypothesis by the prediction that capital negatively Granger-causes nonperforming loans.[6] Because banks with ample capital likely do not face moral hazard incentives, we test the moral hazard hypothesis for a subsample of the data that only includes banks with equity-to-asset ratios below the population median.

---

[5] Moral hazard incentives might occur for a number of reasons, particularly when a bank experiences financial distress. For example, shareholders might exploit creditors by substituting riskier assets for safer ones if creditors do not have sufficient information or time to react. Also, if deposit insurance is imperfectly risk-priced, there may be incentives for banks to undertake risks that will be borne largely by the insurer.

[6] Any number of events may occur after the increase in nonperforming loans (e.g., an increase in operations costs to deal with these loans), but we exclude these possibilities from Figure 1 because they are not needed to identify the moral hazard hypothesis. Furthermore, we do not necessarily predict further erosion of capital after nonperforming loans increase -- the expected return from increasing portfolio risk could be either positive or negative, depending upon whether the additional revenues from higher loan rates are offset by the costs of dealing with the higher expected nonperforming loans.

Note that, unlike the other three theories in Appendix A, the moral hazard hypothesis is not a theory of the relationship between problem loans and measured cost efficiency, but rather of the relationship between problem loans and bank capital ratios. We include it for several reasons. First, moral hazard gives an alternative explanation for nonperforming loans, so the effects of measured cost efficiency on nonperforming loans could be biased if the potential effects of capital were neglected. Second, moral hazard effects can accentuate or magnify the effects of the other three hypotheses, each of which may cause a reduction in the bank's capital ratio. Finally, it is important to include moral hazard because it is a leading theory of problem loans and bank failures with different policy implications from the other hypotheses.

It is important to distinguish between the moral hazard and the skimping hypotheses because they are similar in several respects. Under both hypotheses, banks intentionally take on additional credit risk in their loan portfolios that tends to result in extra nonperforming loans, and in both scenarios this risk-taking may or may not pay off in terms of higher long-run earnings and capital for individual banks. However, the economics of these two hypotheses are quite different. The main purpose of skimping is to save real resources on operating costs, whereas the main purpose of moral hazard is to shift costs to other parties. Skimping creates a social loss only if the total losses associated with the poorly performing loan portfolio exceed the value of the real resources saved by skimping on loan underwriting and monitoring. In contrast, moral hazard behavior is a complete waste from a social viewpoint, since it simply exploits externalities without creating any value.

In an extreme case, all four hypotheses could affect the same bank at the same time. For example, bad luck could befall a poorly managed bank that also happens to be skimping on loan monitoring expenses, and that bank might then respond to moral hazard incentives in an attempt to recapitalize itself. The Granger-causality tests defined in the next section are designed to reveal the industry-wide, or average, inter-temporal relationships among the variables.

Public Policy and Research Implications of the Hypotheses

Each of the four hypotheses has different implications for public policy and future research. The main implications are summarized in Appendix B. The bad luck hypothesis suggests that bank failures are caused primarily by uncontrollable external events, such as a regional downturn or the failure of a large employer in a small bank's lending area. The hypothesis implies that prudential regulation and supervision could reduce the risk of failure by limiting banks' exposures to external shocks (e.g., limits on loan concentrations, allowing interregional diversification through interstate mergers and loan sales, or encouraging low loan-to-asset ratios) or by better insulating banks from external shocks (e.g., requiring high levels of capital). For the purposes of research, the bad luck hypothesis suggests that local or regional economic conditions are important determinants of bank performance. Furthermore, if the bad luck hypothesis dominates the other explanations of the relationship between problem loans and measured cost efficiency, future efficiency measurement should control for nonperforming loans in the cost function. This would help remove by statistical means the extra costs of dealing with nonperforming loans -- which were caused by bad luck, not by managerial inefficiency -- rather than erroneously counting these extra costs as inefficiency as is the case in conventional cost analyses that ignore the effects of problem loans.

In contrast to the bad luck hypothesis, the bad management hypothesis implies that bank failures primarily result from poor managers who allow both costs and loan losses get out of control and deplete the financial capital of the bank. This suggests that bank supervision and bank research should consider cost efficiency along with other traditional predictors of troubled banks such as loan losses and credit risk. Also in contrast to the bad luck hypothesis, the bad management hypothesis implies that cost efficiency measurement should *not* control for nonperforming loans in the cost function. This is because nonperforming loans spring from the same source as the efficiency being measured -- the poor job of management in conducting the affairs of the bank. If the bad management hypothesis is dominant, controlling for nonperforming loans in the cost function artificially increases measured cost efficiency by removing statistically the part of the cost inefficiencies that is correlated with inefficient portfolio management.

The policy and research implications of the skimping hypothesis are different from those of the other hypotheses on several grounds. First, because bank managers who skimp purposely devote fewer inputs to loan underwriting and monitoring, which on average causes nonperforming loans to increase, bank supervisors and researchers studying bank failure might be advised to examine banks' internal credit control procedures (e.g., loan review, collateral appraisal).[7] Second, the skimping hypothesis implies that it may be inappropriate to measure cost efficiency in the short-run, because skimping banks are engaging in long-run strategies. Estimates of short-run cost efficiency will differ depending upon where in Appendix A, panel C the cost efficiency measurement takes place -- skimping banks have relatively low short-run costs when loan quality is good, but may have relatively high short-run costs once nonperforming loans increase. A better strategy may be to estimate long-run cost efficiency using several years of cost data -- years in which skimping saves on costs would be averaged with years in which the nonperforming loans may force up costs. Third, the skimping hypothesis implies that the cost function should *not* control for nonperforming loans, because these performance problems are endogenous to the skimping bank's optimization strategy,

[7] See Udell (1989) for a discussion of internal loan reviews at banks.

which may or may not be more efficient than other strategies. Under the skimping hypothesis, including nonperforming loans in the cost function effectively holds constant an endogenous outcome, and makes skimping banks appear more efficient over the long run by not counting the extra costs that may be created by the skimping strategy. Finally, the skimping hypothesis would also tend to favor estimation of long-run *profit* efficiency in place of either short-run or long-run cost efficiency. Output quality is a choice variable for banks under the skimping hypothesis, and output quality affects on both costs (lower underwriting and monitoring costs) and revenues (higher expected nonperforming loans). Thus, cost minimization and revenue maximization cannot be well separated under this hypothesis.[8] Similar to the arguments for cost efficiency, measurement of profit efficiency should not control for nonperforming loans under the skimping hypothesis because nonperforming loans are an outcome of the bank's optimization strategy.

Under the moral hazard hypothesis, failure can result when banks take additional risks in response to declining financial capital. This hypothesis implies that bank supervisors should monitor capital ratios carefully and require actions to raise the ratios quickly when they become low, similar to the intent of the prompt corrective action feature of the FDIC Improvement Act of 1991. Additional implications of this hypothesis suggest that bank regulators and researchers continue their traditional focus on the effects of capital, how to measure it, and how to better set and enforce capital requirements. Because the moral hazard hypothesis is not a theory of the relationship between problem loans and measured cost efficiency, it contains no implications for whether cost efficiency measurement should control for nonperforming loans.

---

[8] See Berger, Hancock, and Humphrey (1993); DeYoung and Nolle (1995); Akhavein, Swamy, and Taubman (1995); and Humphrey and Pulley (1995) for estimates of profit efficiency for banks.

III. The Econometric Model

We apply Granger-causality techniques to the data on nonperforming loans, measured short-run cost efficiency, equity capital, and some control variables to test which of the four hypotheses -- bad luck, bad management, skimping, and/or moral hazard -- are consistent with the data. The model is sufficiently general that almost any combination of these hypotheses could be supported by the data. Of the four hypotheses, only two are mutually exclusive -- the bad management and skimping hypotheses predict opposite signs for the Granger-causality from measured cost efficiency to nonperforming loans -- and hence cannot be simultaneously supported by the same set of data. Again, we note that Granger-causality tests are gross statistical associations. As with any econometric procedure, application of Granger-causality to the data can only indicate consistency or inconsistency with an hypothesis, not proof of economic causation. However, because our hypotheses each imply unique time-ordered and signed relationships among pairs of the three important variables (nonperforming loans, cost efficiency, and capital), Granger-causality tests may indicate which of our hypotheses are consistent or inconsistent with the data on U.S. commercial banks.

The Granger-causality model is specified as follows:

$$(1)\quad NPL_{i,t} = f_1(NPL_{i,lag}, X\text{-}EFF_{i,lag}, CAP_{i,lag}, RWA_{i,lag}, YEAR_t, REGION_i, YEAR_t \bullet REGION_i) + \epsilon_{1i,t}$$

$$(2)\quad X\text{-}EFF_{i,t} = f_2(NPL_{i,lag}, X\text{-}EFF_{i,lag}, CAP_{i,lag}, RWA_{i,lag}, YEAR_t, REGION_i, YEAR_t \bullet REGION_i) + \epsilon_{2i,t}$$

$$(3)\quad CAP_{i,t} = f_3(NPL_{i,lag}, X\text{-}EFF_{i,lag}, CAP_{i,lag}, RWA_{i,lag}, YEAR_t, REGION_i, YEAR_t \bullet REGION_i) + \epsilon_{3i,t}$$

Definitions and descriptive statistics for all of these variables are shown in Table 1.

Table 1
Definitions and Descriptive Statistics for Variables in the Granger-Causation Model.
(Statistics are for the four-lag model, N = 57,655.)

		mean	std. dev.
NPL	Nonperforming Loan Ratio = dollar value of loans that are either 90 days past-due or are no longer accruing interest, divided by the value of total loans.	0.0168	0.0209
X-EFF	X-Efficiency = short-term cost efficiency, i.e., percent of maximum cost efficiency achieved by bank based on the estimated best-practice cost frontier for the year in question.	0.9224	0.0381
CAP	equity capital ratio = total equity capital divided by gross total assets.	0.0892	0.0292
RWA	risk-weighted asset ratio = estimated ratio of total risk-weighted assets to gross total assets.	0.5661	0.1275
		% of observations	
NE	= 1 if bank is located in Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, or Vermont; = 0 otherwise.	1.49 %	
SW	= 1 if bank is located in Louisiana, Texas, or Oklahoma; = 0 otherwise.	14.53 %	
CA	= 1 if bank is located in California; = 0 otherwise.	2.49 %	
YEAR(89)	= 1 in 1989; = 0 otherwise.	17.60 %	
YEAR(90)	= 1 in 1990; = 0 otherwise.	17.33 %	
YEAR(91)	= 1 in 1991; = 0 otherwise.	17.07 %	
YEAR(92)	= 1 in 1992; = 0 otherwise.	16.69 %	
YEAR(93)	= 1 in 1993; = 0 otherwise.	16.28 %	
YEAR(94)	= 1 in 1994; = 0 otherwise.	15.03 %	

The dependent variable in equation (1) is the nonperforming loan ratio $NPL_{i,t}$. This ratio equals loans that are either past-due at least 90 days or in nonaccrual status divided by total loans at bank i in year t. NPL is the most commonly agreed-upon definition of problem loans in both the research literature and the trade press. NPL also has the benefit of being almost exclusively objectively defined -- it is less subject to regulatory discretion than are other measures of loan quality, such as loan loss provisions and charge-offs.

The dependent variable in equation (2) is estimated cost efficiency $X\text{-}EFF_{i,t}$, which measures the short-term cost X-efficiency of bank i relative to its peers in year t, or how close the bank is to the estimated best-practice cost frontier for all banks in that year. As shown in the next section, we estimate efficiency on a short-term basis and do not control for nonperforming loans in the estimation. We need to observe short-term fluctuations in efficiency in order to test for inter-temporal patterns between efficiency and other variables. We also need to estimate cost efficiency without controlling for nonperforming loans in order to test for relationships between measured efficiency and nonperforming loans. Depending upon which of the four hypotheses are most consistent with the data, this procedure may or may not be the best way to measure cost efficiency.

The third dependent variable is the ratio of equity capital to assets, $CAP_{i,t}$. CAP measures the bank's financial cushion for absorbing portfolio losses. For testing the moral hazard hypothesis in equation (1), CAP is a good measure of how much shareholders have at risk in the portfolio, which largely determines their ability and incentive to shift risk and losses to uninsured creditors and the deposit insurer by investing in risky loans. The capital equation also serves as a check of the bad luck and bad management hypotheses. Both of these hypotheses culminates in reduced capital, so we would expect increases in NPL and/or decreases in X-EFF to Granger-cause reductions in CAP under these hypotheses, although this equation by itself cannot distinguish among the other hypotheses.

The right-hand-side variables in equations (1) - (3) include lagged values of all the dependent variables, as is standard procedure for Granger-causality models. That is, we attempt to determine whether variable Y_i Granger-causes variable Y_j by testing whether the past history of Y_i adds information in predicting Y_j, after taking into account the past history of all the Y's. We use $NPL_{i,lag}$ to denote a vector containing s lags of $NPL_{i,t}$ (i.e., $NPL_{i,t-s}$, where s=1,...,S). The other lag variables are denoted similarly. In our empirical application below, we estimate the model using three, four, and five annual lags of the variables (i.e., s = 3,4,5). Although we report the results of the four-lag model only, the three- and five-lag models produce the same qualitative results.

The model described thus far is equivalent to a vector-autoregressive (VAR) specification in three variables. However, we also include a number of control variables in all three equations. Because NPL, X-EFF, and CAP are often strongly affected by regional or transitory conditions, we need to control in at least a crude fashion for the economic environment of the bank. We therefore specify $YEAR_t$, a set of dummy variables for each year of the data (excluding a base year), which accounts for changes in the macroeconomy and changes in the regulatory treatment of banks over time. We also specify $REGION_i$, a set of three dummies that indicate whether the bank was located in New England, in the Southwest, or in California, because these three geographic regions suffered particularly bad economic downturns relative to the rest of the country during parts of the sample period.[9] The YEAR and REGION variables appear interactively to account for the fact that regional

[9] New England includes Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont. The Southwest includes Louisiana, Oklahoma, and Texas.

differences in economic conditions change over time. Finally, we specify $RWA_{i,t}$ as the ratio of risk-weighted assets (according to the Basle Accord) to gross total assets for bank i in year t, and include lagged values of RWA in all three equations. RWA provides a crude control for credit risk, and at the very least includes more information than do simple loan-to-asset ratios.[10] We include RWA because certain portfolio mixes will naturally yield more nonperforming loans, because low measured cost efficiency may reflect the costs associated with a more loan-intensive balance sheet, and because capital markets and bank regulators may require banks with certain lending strategies to have higher capital ratios.[11]

Equations (1) - (3) are estimated separately using OLS techniques and a panel of data over T years (t = 1,...,T) and N banks (i = 1,...,N), with specifications of 3, 4, and 5 annual lags (S = 3, 4, or 5). The terms $\varepsilon_{1,i,t}$, $\varepsilon_{2,i,t}$, and $\varepsilon_{3,i,t}$ are independent random error terms. There is no serious concern about the correlations among the random error terms either within an equation or across the three equations, because of the specification of the right-hand-side variables. Including three to five years of lags of the *dependent* variables in each equation should absorb the bulk of any serial correlation in the three error terms. Similarly, including the YEAR, REGION, and YEAR*REGION variables absorbs any cross-section correlation that is constant across banks within a year, across banks within a region, or across banks in a region in a particular year. Also, because the same right-hand-side

---

[10] RWA is designed to capture credit risk taking into account a number of factors such as the type of borrower and the existence of collateral and guarantees. The risk weights have been shown to be positively related to the probability of failure and to some accounting and market measures of risk, although the correspondence has been shown to be relatively weak (see Avery and Berger 1991; Cordell and King 1995; Jones and King 1995).

[11] Including lagged values of RWA on the right-hand-side of equation (1) does not control for moral hazard behavior. When a bank responds to moral hazard incentives by taking increased risks, these risks are likely taken in secret by substituting higher risk loans for lower risk loans *within* a risk-weighted asset loan category, which is not reflected in RWA.

variables are specified for each of the three equations, there is no need to take account of the correlations across the error terms in the different equations. In such a specification, OLS and GLS yield identical estimators.

The key empirical implications of the four hypotheses are easily summarized in terms of the coefficients of equations (1) - (3). For example, the main identifying prediction of the bad luck hypothesis is a negative sum of the coefficients on the NPL lags in the X-EFF equation (2). That is, the bad luck hypothesis predicts that measured cost efficiency will be lower in the several years after high nonperforming loans are reported as the bank has to expend resources dealing with negotiations, defaults, disposing of collateral, etc. We focus on the sum of the lag coefficients, rather than the individual lag coefficients, because the sum measures the total effect of past NPL on current X-EFF, and because the sum is more accurately measured than its individual parts due to collinearity among the individual lagged values. Unlike most Granger-causality exercises, we are most interested in whether the sum of the coefficients is positive or negative, not whether the individual coefficients are all equal to zero. Note that including the other right-hand-side variables in the equation makes our tests more stringent, because (in the case of equation (2)) we are controlling for any predictive effects of lagged cost efficiency, lagged capital, lagged risk-weighted assets, year, and region on current measured efficiency.

Similarly, the bad management hypothesis predicts that the sum of the coefficients on the X-EFF lags in the NPL equation (1) will be negative. The effects of subpar management on operations should be observed in higher costs rather quickly, whereas the nonperforming loan data will reflect the impact of poor management on the loan portfolio only after borrowers have fallen behind in their loan repayments and after the bank has recognized and actually reported this phenomenon. Under the skimping hypothesis, the sum of the coefficients on the X-EFF lags in the NPL equation (1) will be positive, as the banks that improved their measured short-run cost efficiency in the past by intentionally reducing expenditures on loan underwriting and monitoring develop more problem loans in the present. Since the bad management and skimping hypotheses predict opposite signs for the sum of the lagged X-EFF coefficients in the NPL equation, we view this coefficient sum as the net effect of these two opposing theories. The final hypothesis, moral hazard, predicts that the sum of the coefficients on the CAP lags in the NPL equation (1) will be negative, as banks with low capital in the past took increased risks in response to incentives created by low capital, which may result in high levels of problem loans in the present. We test the moral hazard hypothesis only for a subsample of thinly capitalized banks, because banks with high levels of capital do not face moral hazard incentives.

Finally, although all of the empirical implications that uniquely distinguish among the four hypotheses are contained in equations (1) and (2), we have included the capital equation (3) because it is important to complete the overall model and verify the basic relationships that are being investigated. The bad luck and bad management hypotheses all predict that either nonperforming loans, cost efficiency, or both Granger-cause equity capital. In equation (3) we expect to find a negative sum of coefficients on the NPL lags and a positive sum of coefficients on the X-EFF lags, although we cannot distinguish among these hypotheses using this equation. Such findings would extend the stylized facts from the literature that problem loans and inefficiencies are strongly associated with losses of bank capital.

IV. The Cost Efficiency Model

We used the econometric frontier approach (EFA) to generate an annual estimate of cost efficiency for each bank i in each year t.[12] In the EFA method, a frontier cost function is estimated using a statistical procedure that decomposes the error term into two parts. One part captures random disturbances and is assumed to follow a symmetric normal distribution and the other part is assumed to capture cost inefficiency.[13] A one-sided distribution (usually the half normal) is imposed on the cost inefficiency portion of the error term, since inefficiency cannot be negative. We impose a truncated normal distribution on the inefficiency term, which is shown by Stevenson (1980) to be more general and flexible than the typical assumption of a half normal distribution.[14]

We specify the Fourier-flexible functional form for the cost function. This functional form combines a standard translog functional form with the non-parametric Fourier functional form. Because the translog form is only a local approximation, it may perform poorly for observations far from the sample means. In contrast, the Fourier-flexible form adds trigonometric transformations of the variables to globally approximate the underlying cost function over the entire range of data. Mitchell and Onvural (1992), McAllister and McManus (1993), Berger, Leusner, and Mingo (1994), and Berger, Cummins, and Weiss (1995) all found that the Fourier-flexible form dominates the translog.

---

[12] See Bauer (1990) for a review of EFA methods. We cannot use the distribution-free approach because it only generates long-run estimates of cost efficiency. We cannot use the thick cost frontier approach because it estimates cost efficiency only for groups of banks. We do not use data envelopment analysis because it is often difficult to compare the efficiency of individual institutions using this approach. Berger, Hunter, and Timme (1993) discussed the relative strengths and weaknesses of these approaches.

[13] As estimated here, cost inefficiency includes both technical and allocative inefficiencies. These inefficiencies may be the result of numerous managerial (and other) phenomena, including expense preference behavior, agency problems, managerial incompetence, and the degree of managerial effort.

[14] The half normal distribution, which is a special case of the truncated normal distribution, restricts the probability density of the cost inefficiency term to decrease monotonically across banks, whereas the truncated normal does not.

We estimated the following operating cost function for each year t in our analysis:[15]

$$
\begin{aligned}
lnOC \;=\; & \alpha_0 + \sum_{j}^{5} \beta_j lnY_j + \frac{1}{2}\sum_{j}^{5}\sum_{k}^{5} \beta_{jk} lnY_j lnY_k + \sum_{m}^{2} \gamma_m lnW_m \\
& + \frac{1}{2}\sum_{m}^{2}\sum_{n}^{2} \gamma_{mn} lnW_m lnW_n + \sum_{j}^{5}\sum_{m}^{2} \rho_{jm} lnY_j lnW_m \\
& + \sum_{j=1}^{7} [\delta_j \cos Z_j + \theta_j sinZ_j] + \sum_{j=1}^{7}\sum_{k=j}^{7} [\delta_{jk}\cos(Z_j + Z_k) + \theta_{jk}\sin(Z_j + Z_k)] \\
& + \sum_{j=1}^{7}\sum_{k=j}^{7}\sum_{l=k}^{7} [\delta_{jkl}\cos(Z_j + Z_k + Z_l) + \theta_{jkl}\sin(Z_j + Z_k + Z_l)] \\
& + \lambda_L LIMIT + \lambda_U UNIT + \eta \qquad (4)
\end{aligned}
$$

where the subscript that identifies individual banks has been dropped for simplicity. OC is operating (noninterest) expense; Y is a vector of outputs including commercial loans, consumer loans, real estate loans, transactions deposits, and fee-based income;[16] W is a vector of input prices including the prices of labor and physical capital;[17] LIMIT and UNIT are, respectively, dummy variables equal to 1 for banks in states that restricted or banned branch banking during the sample

[15] We estimated equation (4) using maximum likelihood techniques, and imposed the standard symmetry and homogeneity restrictions on the translog portion of the model. Factor share equations were omitted because application of the usual cross-equation restrictions would impose the assumption that the given input proportions were the allocatively efficient ones (see Berger 1993, p. 266).

[16] Transactions deposits include demand deposits, NOW accounts, automatic transfer service accounts, and telephone and pre-authorized transfer accounts. Fee-based income equals gross noninterest income less both service charges on deposit accounts and gains(losses) from securities and foreign exchange trading. Fee-based income is included to control for activities other than deposits and loans such as off-balance sheet activities and trust services.

[17] The price of labor equals salaries and benefits divided by the number of full time equivalent workers. The price of physical capital equals expenditures on equipment and premises divided by the book value of physical assets.

period. The error term η is a composite expression: $\eta = \ln U + \ln V$, where lnU captures cost inefficiency and is distributed as a truncated normal variable, and lnV captures random error and is distributed as a normal variable. The Z_j values are functions that rescale the $\ln Y_j$ and the $\ln W_m$ so that they fall on the interval $[.1*2\pi , .9*2\pi]$.[18]

We estimate an operating cost function (i.e., noninterest expense only) because operating costs have been shown elsewhere to comprise the bulk of excess costs at banks (see Berger and Humphrey (1991)); because the costs associated with credit evaluation, loan monitoring, and administering to problem loans are mainly noninterest expenses; and because interest cost inefficiencies are substantially affected by interest rate levels, which vary considerably during our sample period. As discussed above, we do not control for loan quality in the cost function, because we do not want to remove from our estimates of cost efficiency the expenses that are correlated with inefficient portfolio management.[19] We estimate 10 separate annual cost frontiers, rather than estimating a common frontier across time, because conditions change from year-to-year -- that is, the technology or bank that is most efficient in one year may not be the most efficient in another year.

We construct the cost efficiency variable X-EFF as follows. Using equation (4), we estimate the expected value of the inefficiency term lnU conditional on η, due to Jondrow, Lovell, Materov, and Schmidt (1977). We then transform the estimates of lnU so that they vary positively with cost efficiency, rather than with cost inefficiency. For bank i in year t, $\text{X-EFF}_{i,t}$ equals

---

[18] See Berger, Leusner, and Mingo (1994) for a derivation of, and a justification for, this truncation. Because of computational limitations, we restricted $j = k = l$ in the triple summation terms.

[19] We also do not include either financial capital or the price of borrowed funds in the cost function, because our cost measure excludes interest expense.

U_{min}/U_i, the estimated value of U for the most efficient bank divided by the estimated value of U for bank i, where $U = e^{\ln U}$. Thus, X-EFF is positive, increases with cost efficiency, and has an upper bound of one.

V. Data

The data set consists of an unbalanced panel of annual observations of U.S. commercial banks from 1985 through 1994. Some years have a greater number of observations than others because of failure, acquisition, or entry during the sample period. We exclude observations for which at least five years worth of lagged variables could not be observed. The size of the panel that we actually used in the estimations depended on the number of lags included in the model -- there were 69,742 observations, 57,780 observations, and 46,504 observations, respectively, when s = 3, 4, and 5 lags. Table 1 shows definitions and sample means and standard deviations for the variables used in the four-lag Granger-causality model.

All variables were constructed using end-of-year data from the annual Reports of Condition and Income (call reports) for 1985 through 1994, with the exception of UNIT and LIMIT, which were constructed from information in Amel (1993). We use annual data because end-of-year call report data are more accurate than quarterly call report data and to reduce the amount of random error in the cost efficiency estimation. By using annual data, however, we may understate the magnitude of any Granger causation among NPL, X-EFF, and CAP if a substantial amount of the inter-temporal effects associated with our four hypotheses occur within a year.

Operating costs (OC) and the numerators used to construct the input prices (W) are flow variables that reflect accumulated activity over the course of each year. The nonperforming loan ratio (NPL), the capital ratio (CAP), the risk-weighted asset ratio (RWA), the cost function output vector (Y), and the denominators used to construct input prices are stock variables, and are included here as averages of beginning-of-year and end-of-year values. Because the natural log of zero equals -∞, a small positive amount ($1,000) was added to all output variables for all banks. A small number of observations in each year were excluded from the efficiency estimations because either assets or total costs were reported as zero, because input prices could not be constructed, or because constructed input prices were unrealistically large or small.[20]

VI. Results

Efficiency Estimation Results

Means and standard deviations for the annual estimates of cost efficiency are displayed in Table 2. The average bank is between 90 percent and 95 percent efficient over the sample period, a higher level of cost efficiency than is found in most other econometric frontier (EFA) studies of commercial banks. The likely reason for this difference is our more general and flexible specification. Earlier EFA studies may have understated cost efficiency by imposing too much structure on both the cost function (by using the translog form) and on the inefficiency term (by using the half normal distribution).

To examine this issue, we estimated X-EFF for three additional variants of the cost model. A comparison of the results using data from 1994 is shown in Table 3.[21] Variant A shows the 1994 values using the Fourier-flexible form and specifies the truncated normal distribution for lnU; variant B uses the Fourier-flexible form but imposes a half normal distribution on lnU; variant C uses a

---

[20] This also caused even more observations to be deleted from the Granger-causality tests because five lags of efficiency were needed for these tests.

[21] Tests using data from other years during the sample period yielded similar results.

Table 2

Estimated X-Efficiency, 1985-1994

	N	mean	std. dev.
1985	12,606	0.9180	0.0460
1986	12,343	0.9402	0.0217
1987	11,933	0.9499	0.0195
1988	11,546	0.9086	0.0476
1989	11,278	0.9266	0.0345
1990	11,027	0.9071	0.0456
1991	10,753	0.9123	0.0522
1992	10,452	0.9122	0.0425
1993	10,130	0.9290	0.0318
1994	9,622	0.9466	0.0178

Table 3

Estimates of X-Efficiency, 1994

Various functional forms and distributional assumptions.
** and * indicate significantly different from the mean of (A) at the 1 percent and 5 percent levels.

		mean	std. dev.	rank order correlation with (A)
A.	Fourier-flexible, truncated normal	0.9466	0.0178	
B.	Fourier-flexible, half normal	0.9191**	0.0312	0.9999
C.	Translog, truncated normal	0.8989**	0.0411	0.9688
D.	Translog, half normal	0.8780**	0.0517	0.9690

standard translog form (i.e., drops the sine and cosine terms) and specifies a truncated normal distribution for lnU; and variant D uses the translog form and imposes a half normal distribution on lnU. Thus, Variant A is the model used throughout this paper and is the least restrictive, variant D is the model generally used elsewhere and is the most restrictive, and variants B and C lie between these extremes. Estimated X-efficiency is significantly lower when the translog is substituted for the Fourier-flexible, and is also significantly lower when the half normal distribution is substituted for the truncated normal distribution. Hence, the structure most often employed in the literature (variant D) may misdiagnose some of the variability in the cost data as inefficiency, and shows about twice as much inefficiency as the most general form used her (variant A). Nevertheless, the rank efficiency ordering of the banks is very similar across the four variants. Finally, neither the functional form of the cost function nor the distribution imposed on the inefficiency terms substantially affect the rank efficiency ordering of the banks in our sample.[22, 23]

Granger-Causality Test Results

Table 4 displays the coefficient estimates from equations (1)-(3) using four lags on the right-hand-side variables. All of the major results were robust to the number of lags included in the model -- results using three and five lags are not shown here in order to conserve space.

In the X-EFF equation, we find evidence in support of the bad luck hypothesis. The sum of

[22] These results stand in contrast to Hunter and Timme (1993), who found that the rank efficiency ordering of banks is sensitive to how the cost function is specified, and also to Bauer, Berger, and Humphrey (1993), who found that the rank efficiency ordering of banks varies with efficiency approach.

[23] We also directly tested the restrictions implied by a translog specification. That is, we tested whether the coefficients on all of the trigonometric terms in the Fourier-flexible form were jointly equal to zero. The F-statistic implied a rejection of the translog null hypothesis at the 1 percent level for both the half normal and the truncated normal specifications.

Table 4
Granger-Causality Tests
Four lags.

	(1) NPL	(2) X-EFF	(3) CAP
INTERCEPT	.0011	.2266**	-.0032*
	(0.61)	(83.92)	(2.43)
NPL(-1)	.5626**	-.1316**	-.0146**
	(140.45)	(22.99)	(5.19)
NPL(-2)	.0593**	.0099	.0142**
	(13.56)	(1.58)	(4.61)
NPL(-3)	.0084*	.0351**	.0092**
	(2.09)	(6.13)	(3.27)
NPL(-4)	.0134**	.0267**	.0075**
	(4.18)	(5.84)	(3.33)
NPL(total)	.6437**	-.0599**	.0163**
	(183.31)	(11.94)	(6.70)
X-EFF(-1)	-.0164**	.5940**	.0070**
	(6.34)	(160.98)	(3.89)
X-EFF(-2)	.0056	.1809**	-.0022
	(1.83)	(41.50)	(1.06)
X-EFF(-3)	.0062	-.0162**	.0045*
	(1.94)	(3.54)	(2.02)
X-EFF(-4)	-.0022	.0032	.0055**
	(0.84)	(0.86)	(3.00)
X-EFF(total)	-.0068**	.7619**	.0148**
	(3.35)	(263.04)	(10.44)
CAP(-1)	-.0540**	-.0538**	.9358**
	(9.21)	(6.41)	(227.47)
CAP(-2)	.0418**	.0521**	-.0225**
	(5.11)	(4.46)	(3.92)
CAP(-3)	.0305**	.0119	.0131*
	(3.67)	(1.01)	(2.24)
CAP(-4)	.0011	-.0056	.0074
	(0.19)	(0.65)	(1.74)
CAP(total)	.0194**	.0046	.9338**
	(7.63)	(1.27)	(521.84)
RWA(-1)	.0102**	.0178**	-.0184**
	(8.21)	(10.05)	(21.15)
RWA(-2)	.0057**	-.0171**	.0063**
	(3.51)	(7.38)	(5.51)
RWA(-3)	.0016	-.0032	.0017
	(1.02)	(1.40)	(1.53)
RWA(-4)	-.0035**	.0014	.0028**
	(2.96)	(0.85)	(3.43)
RWA(total)	.0140**	-.0011	-.0076**
	(23.95)	(1.29)	(18.14)

Absolute values of t-statistics are in parentheses ( ).

** and * indicate significance at the 1 percent and 5 percent levels, respectively.

Table 4 (continued)

	(1) NPL	(2) X-EFF	(3) CAP
YEAR(90)	.0018**	-.0207**	-.0019**
	(6.39)	(51.41)	(9.48)
YEAR(91)	.0012**	-.0077**	.0007**
	(4.15)	(19.49)	(3.77)
YEAR(92)	-.0018**	-.0065**	.0020**
	(6.76)	(17.18)	(10.94)
YEAR(93)	-.0020**	.0080**	.0033**
	(7.28)	(20.21)	(17.05)
YEAR(94)	-.0024**	.0147**	-.0002
	(8.38)	(36.59)	(1.23)
NE	.0155**	.0019	-.0015*
	(13.98)	(1.20)	(1.96)
NE*YEAR(90)	.0049**	.0036	-.0022
	(3.06)	(1.58)	(1.90)
NE*YEAR(91)	-.0023	-.0098**	-.0046**
	(1.36)	(4.12)	(3.94)
NE*YEAR(92)	-.0136**	-.0064**	.0014
	(7.91)	(2.63)	(1.14)
NE*YEAR(93)	-.0115**	-.0016	.0024*
	(6.56)	(0.63)	(1.99)
NE*YEAR(94)	-.0126**	.0003	.0027*
	(7.02)	(0.12)	(2.12)
SW	.0084**	.0056**	-.0036**
	(19.16)	(8.99)	(11.77)
SW*YEAR(90)	-.0056**	-.0088**	.0009*
	(9.14)	(10.05)	(2.05)
SW*YEAR(91)	-.0063**	-.0090**	.0028**
	(10.36)	(10.30)	(6.50)
SW*YEAR(92)	-.0063**	-.0088**	.0043**
	(10.27)	(10.07)	(10.03)
SW*YEAR(93)	-.0070**	-.0044**	.0046**
	(11.33)	(5.04)	(10.54)
SW*YEAR(94)	-.0070**	-.0022*	.0029**
	(11.15)	(2.48)	(6.50)
CA	-.0023*	.0065**	.0024**
	(2.07)	(4.18)	(3.11)
CA*YEAR(90)	.0013	.0001	.0033**
	(0.83)	(0.05)	(3.07)
CA*YEAR(91)	.0139**	-.0098**	-.0034**
	(9.44)	(4.66)	(3.34)
CA*YEAR(92)	.0146**	-.0126**	-.0026*
	(10.14)	(6.10)	(2.56)
CA*YEAR(93)	.0157**	-.0128**	-.0053**
	(11.03)	(6.26)	(5.34)
CA*YEAR(94)	.0071**	-.0033	.0001
	(4.96)	(1.63)	(0.11)
Adjusted R^2	.4944	.6892	.8722
N	57,655	57,655	57,655

Absolute values of t-statistics are in parentheses ( ).
** and * indicate significance at the 1 percent and 5 percent levels, respectively.

the lagged NPL coefficients is -.0599 and is significant at the 1 percent level. Thus, the Granger-causality results are consistent with the notion that after loans become past due or nonaccruing, operating costs rise because of the difficulty in dealing with these loans, as predicted by the bad luck hypothesis. However, the economic impact of this result was small -- for the average bank, a one standard deviation increase in NPL (from .0168 to .0377) predicts a cumulative reduction in measured cost efficiency over four years from .9224 to .9211, or a 1.7 percent increase in average inefficiency.[24] Note that the lagged coefficient sums on CAP and RWA were not significantly different from zero in this equation.

The results of the NPL equation suggest that the bad management hypothesis dominates the skimping hypothesis. The sum of the coefficients on the lagged X-EFF variables equals -.0068 and is significant at the 1 percent level. Thus, Granger-causality suggests that after measured cost efficiency declines, nonperforming loans increase because of poor loan portfolio management, as predicted by the bad management hypothesis. The economic impact of this result is also small -- for the average bank, a one standard deviation reduction in X-EFF (from .9224 to .8843) predicts a cumulative increase in the nonperforming loan ratio over four years from .0168 to only .0171, or a 1.8 percent increase in nonperforming loans. As expected, the lagged RWA coefficient sums are positive and significant in the NPL equation. This result suggests that a relatively risky loan portfolio mix will eventually yield relatively high numbers of nonperforming loans.

The moral hazard hypothesis prdicts that reductions in capital lead to increases in

---

[24] These calculations reflect only the direct effect of changes in lagged Y_i on current Y_j because these direct effects are the most appropriate measures for the hypotheses we are testing. We exclude indirect effects -- such as the effect of lagged Y_i on current Y_j through other lagged values of the Y's -- from our calculations in order to avoid mixing the inter-temporal effects associated with our four different hypotheses.

nonperforming loans, at least for thinly capitalized banks, because these banks are most likely to face moral hazard incentives. Perhaps surprisingly, when the NPL equation is estimated for the entire population of banks the sums of the lagged CAP coefficients are positive and significant. However, this is not an appropriate test of the moral hazard hypothesis, because banks with high or adequate levels of capital likely do not face moral hazard incentives, we test the moral hazard hypothesis by estimating the NPL equation separately using only the annual observations for which CAP was below the population median.[25] Partial results of these regressions are shown in the first panel in Table 5. The sum of the coefficients on the lagged CAP variables equals -.0510 and is significant at the 1 percent level. These data support the moral hazard hypothesis, and suggest that, on average, thinly capitalized banks take increased portfolio risk, which results in higher levels of problem loans in the future. For the average low-capital bank, a one standard deviation reduction in CAP (from .0712 to .0578) predicts a cumulative increase in the nonperforming loan ratio over four years from .0186 to .0193, or a 3.8 percent increase in nonperforming loans. Although the overall impact of this result is small, it may be quite large for some individual banks, given that our results are effectively an average of the behavior of banks that do, and banks that do not, respond to moral hazard incentives.

As stated above, the CAP equation does not allow us to distinguish between any of our four major hypotheses, but is included for completeness because three of the four hypotheses culminate in a loss of capital. The sum of the coefficients on the lagged X-EFF variables was positive, evidence that cost-inefficient banks are likely to have low and perhaps even negative earnings, which

[25] We selected the subsample based on the median values of CAP(-1).

Table 5
Granger-Causality Tests
Partial Results for Subsample Estimations. Four lags.

	coefficient estimates	t-statistics (absolute values)
Dependent variable is NPL		
Banks with CAP(-1) less than population median in individual years (N = 28,826, adjusted R^2 = .5103).		
CAP(-1)	-.1575**	(11.99)
CAP(-2)	.0855**	(6.20)
CAP(-3)	.0084	(0.65)
CAP(-4)	.0126	(1.35)
CAP(total)	-.0510**	(4.86)
Dependent variable is CAP		
Banks with CAP(-1) less than population median in individual years (N = 28,826, adjusted R^2 = .5039).		
NPL(-1)	-.0302**	(8.96)
NPL(-2)	.0160**	(4.32)
NPL(-3)	.0110**	(3.20)
NPL(-4)	.0112**	(4.07)
NPL(total)	.0080*	(2.58)
Dependent variable is CAP		
Banks with CAP(-1) greater than population median in individual years (N = 28,826, adjusted R^2 = .8557).		
NPL(-1)	-.0068	(1.45)
NPL(-2)	.0067	(1.34)
NPL(-3)	.0024	(0.53)
NPL(-4)	-.0002	(0.05)
NPL(total)	.0021	(0.52)
Dependent variable is NPL		
Banks with X-EFF(-1) greater than population median in every year (N = 12,756, adjusted R^2 = .4943).		
X-EFF(-1)	.0002	(0.02)
X-EFF(-2)	.0117	(1.00)
X-EFF(-3)	.0133	(1.17)
X-EFF(-4)	-.0019	(0.25)
X-EFF(total)	.0233**	(3.13)

** and * indicate significance at the 1 percent and 5 percent levels, respectively.

Granger-cause reductions in capital. The sum of the lagged RWA coefficients is negative and significant, evidence that higher risk lending strategies also Granger-cause reductions in capital on average. However, the sum of the lagged NPL coefficients is positive and significant, a counter-intuitive result suggesting that high levels of nonperforming loans Granger-cause high capital ratios. It is possible that banks with high and low levels of capital behave differently in response to increases in problem loans. In particular, thinly capitalized banks may be under pressure from bank regulators, uninsured creditors, and the capital markets to improve their marginal capital positions, while banks with large capital cushions can afford to book loan losses without altering their lending or capital strategies.[26] To investigate this possibility, we estimated the CAP equation using only the annual observations for which CAP was above the population median, and then once again using only the annual observations for which CAP was below the population median.[27] Partial results of these regressions are shown in the second and third panels in Table 5. We continue to find that nonperforming loans positively Granger-cause capital ratios at the low-capital banks, which suggests that these banks take action to replenish capital after nonperforming loans increase, perhaps under pressure from regulators and markets. This result disappears at the high-capital banks for which we find no Granger-causality between nonperfoming loans and capital.

Although we find no evidence in support of the skimping hypothesis in Table 4, these results do not preclude the possibility of skimping in individual banks. Our Granger-causality results are average relationships, so it is possible that the bad management relationship that exists in the

---

[26] A survivor bias in our data set may contribute to this result. Because we use end-of-year data we only observe the low-capital banks that survived the year, not the low-capital banks that failed during that year. The expected negative relationship between increasing problem loans and reductions in capital is likely to be strongest in failing banks, since these banks essentially lost all of their capital.

[27] We selected these subsamples based on the median values of CAP(-1).

majority of banks simply dominates the skimping relationship that exists in a few banks. We might expect to find concentrations of "skimpers" among the most cost-efficient banks, i.e., banks that consistently skimp on loan underwriting and monitoring in order to cut costs, but manage the resulting loan quality problems in a cost effective fashion. To investigate this possibility, we estimated the NPL equation for the subsample of banks that were consistently cost efficient, having efficiency greater than the median in *every* year of our analysis.[28] Partial results of these regressions are shown in the fourth panel in Table 5. In the NPL equation, the sum of the lagged X-EFF coefficients equals .0233 and is significant at the 1 percent level, suggesting that measured cost efficiency positively Granger-causes nonperforming loans among highly efficient banks, supporting the skimping hypothesis for this subset of banks. For the average bank in this subsample, a one standard deviation increase in X-EFF (from .9522 to .9675) predicts a cumulative increase in the nonperforming loan ratio over four years from .0128 to .0132, or a 3.1 percent increase in nonperforming loans. To the extent that skimping behavior did occur, it is not at all clear that these banks were materially harmed -- all of the banks in this subsample managed to remain among the most cost-efficient banks in the industry for the entire sample period. These results may reflect optimizing behavior among the most cost-efficient banks in the industry -- banks that, at the margin, willingly accepted the marginal increases in loan problems in exchange for marginal cuts in underwriting and monitoring expenses.

VII. Conclusions

[28] We selected the subsample based on the median values of X-EFF(-1). We required banks to be relatively cost efficient in every year, rather than just in individual years, to avoid including banks that only appeared to be efficient in individual years because they were skimping. This sampling procedure ensures that we observed only long-run cost efficient banks -- long-run cost-efficient skimpers as well as long-run cost-efficient non-skimpers.

Loan quality at financial institutions has received close scrutiny since the failure waves in the commercial banking and thrift industries. Simultaneous with these developments, the empirical literature on productive efficiency in financial institutions has burgeoned. In this paper, we weave together these two strands of research. We employ Granger-causality techniques to test whether cost efficiency in banks pre-dates loan quality, whether loan quality pre-dates cost efficiency, or both, using pooled cross section-time series data on nonperforming loans, operating cost efficiency, equity capital ratios, and other variables for U.S. commercial banks between 1985 and 1994.

We test three hypotheses regarding the intertemporal relationships between loan quality and efficiency. The "bad luck" hypothesis posits that exogenous events can cause nonperforming loans to increase, and that after time passes the extra expenses associated with these loans (e.g., more intensive monitoring, negotiating workout arrangements, handling collateral if default occurs) will be reflected in lower measured cost efficiency. Hence, nonperforming loans should negatively Granger-cause cost efficiency. The "bad management" hypothesis posits that poorly run banks do bad jobs at both cost control and at loan underwriting and monitoring, and that after time passes this slack leads to increases in problem loans as borrowers fall behind on their loan repayments. Hence, low cost efficiency should precede increases in nonperforming loans, i.e., cost efficiency should negatively Granger-cause nonperforming loans. The "skimping" hypothesis posits that banks might achieve low costs by under-spending on loan underwriting and monitoring in the short run, and after time passes this slack results in increases in problem loans. Hence, measured cost efficiency should positively Granger-cause nonperforming loans. We also examine the familiar "moral hazard" hypothesis by testing whether equity capital negatively Granger-causes nonperforming loans.

Our results suggest that the inter-temporal relationships between loan quality and cost efficiency run in both directions. The data provide support for the bad luck hypothesis. Increases in nonperforming loans tend to be followed by decreases in measured cost efficiency, suggesting that problem loans cause banks to increase spending on monitoring, working out, and/or selling off problem loans. For the industry as a whole, the data favor the bad management hypothesis over the skimping hypothesis -- decreases in measured cost efficiency are generally followed by increases in nonperforming loans, evidence that bad management practices are manifested not only in excess expenditures, but also in subpar underwriting and monitoring practices that eventually lead to nonperforming loans. For a subset of banks that are consistently efficient, however, increases in measured cost efficiency precede increases in nonperforming loans, consistent with the skimping hypothesis that banks trade short-run expense reductions for long-run reductions in loan quality. Finally, decreases in bank capital ratios precede increases in nonperforming loans for banks with low capital ratios, evidence that thinly capitalized banks may respond to moral hazard incentives by taking increased portfolio risks.

Depending on whether or not these results are confirmed by future research, our findings may have research and policy implications. The bad luck hypothesis implies that the major risks facing financial institutions are exogenous events, and as such supports policies that limit the exposure of institutions to external shocks and/or insulate institutions from the effects of shocks. In contrast, the bad management hypothesis implies that the major risks facing financial institutions are caused internally, and implies that cost efficiency be included as an indicator of potentially troubled banks.

The empirical results also have implications for estimating the efficiency of financial institutions. The empirical support for the bad luck hypothesis suggests that controls for loan quality may be needed to avoid understating cost efficiency at unlucky banks, while the evidence supporting the bad management hypothesis implies that controlling for loan quality may overstate cost efficiency at poorly run banks. Neither hypothesis clearly dominates the other. The skimping hypothesis, which received much less empirical support, also suggests that controls for loan quality be excluded when estimating efficiency. Ultimately, whether or not one controls for loan quality should rest on the particular efficiency application at hand. We also find that measured cost efficiency varies substantially with the degree of flexibility in the specification of the cost equation and the error structure. Measured cost efficiency is lower when the specification is less flexible, perhaps because some of the misspecification is measured as cost inefficiency.

The skimping hypothesis also implies that efficiency estimation should use several years of cost or profit data, so that years in which skimping saves on costs would be averaged with years in which the resulting nonperforming loans force costs up and revenues down. Similarly, profit efficiency is preferred over cost efficiency under the skimping hypothesis. This is because both costs and revenues are affected by the decision to skimp.

We stress the limitations of our analysis. The inter-temporal relationships revealed by Granger-causality techniques are gross statistical associations only, and do not necessarily prove economic causation. However, these relationships are indicative of which among the alternative hypotheses are consistent with the data. Future research might use other statistical techniques to reveal the inter-temporal relationships between loan quality and productive efficiency in financial institutions; attempt to decompose the determinants of loan quality into internal versus exogenous factors; or focus on the empirical consequences of controlling for loan quality when estimating efficiency.

References

Akhavein, Jalal D., P.A.V.B. Swamy, and Stephen B. Taubman, "A General Method for Deriving the Efficiencies of Banks from a Profit Function," Board of Governors of the Federal Reserve System, working paper, 1995.

Amel, Dean, "State Laws Affecting Commercial Bank Branching, Multibank Holding Company Expansion, and Interstate Banking," Board of Governors of the Federal Reserve System, working paper, 1993.

Avery, Robert B., and Allen N. Berger, "Risk-Based Capital and Deposit Insurance Reform," *Journal of Banking and Finance* 15: 847-874, 1991.

Barr, Richard, and Thomas Siems, "Predicting Bank Failure Using DEA to Quantify Management Quality," Federal Reserve Bank of Dallas, Financial Industry Studies Working Paper No. 1-94, January 1994.

Bauer, Paul W., "Recent Developments in the Econometric Estimation of Frontiers," *Journal of Econometrics*, 46: 39-56, 1990.

Bauer, Paul W., Allen N. Berger, and David B. Humphrey, "Efficiency and Productivity Growth in U.S. Banking," in H.O. Fried, C.A.K. Lovell, and S.S. Schmidt, eds., *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford: Oxford University Press: 386-413, 1993.

Becher, David, Robert DeYoung, and Thomas Lutton, "Projecting Resolved Assets in Banks: A Comparison of Different Methods," Office of the Comptroller of the Currency, working paper, 1995.

Berg, Sigbjørn, Finn Førsund, and Eilev Jansen, "Malmquist Indices of Productivity Growth during the Deregulation of Norwegian Banking, 1980-89," *Scandinavian Journal of Economics* 94, 1992.

Berger, Allen N., "Distribution-Free Estimates of Efficiency in the U.S. Banking Industry and Tests of the Standard Distributional Assumptions," *Journal of Productivity Analysis* 4: 261-292, 1993.

Berger, Allen N., "The Relationship between Capital and Earnings in Banking," *Journal of Money, Credit, and Banking* 27: 432-56, 1995.

Berger, Allen N., Cummins, J. David, and Mary A. Weiss, "The Coexistence of Multiple Distribution Systems for Financial Services: The Case of Property-Liability Insurance," working paper, 1995.

Berger, Allen N., Diana Hancock, and David B. Humphrey, "Bank Efficiency Derived from the Profit Function," *Journal of Banking and Finance*, 17: 317-347, 1993.

Berger, Allen N., and David B. Humphrey, "Measurement and Efficiency Issues in Commercial Banking," in Z. Griliches, ed., *Output Measurement in the Service Sectors*, National Bureau of Economic Research, Studies in Income and Wealth, Vol. 56, University of Chicago Press (Chicago, IL): 245-79, 1992.

Berger, Allen N., William Hunter, and Stephen Timme, "The Efficiency of Financial Institutions: A Review and Preview of Research Past, Present, and Future,"*Journal of Banking and Finance*, 17: 221-249, 1993.

Berger, Allen N., John H. Leusner, and John J. Mingo, "The Efficiency of Bank Branches," Board of Governors of the Federal Reserve, Finance and Economics Discussion Series Paper 94-26, 1994.

Board of Governors of the Federal Reserve System, *Federal Reserve Bulletin*, Washington, D.C., volumes 73 (1987) through 79 (1993).

Cordell, Lawrence R., and Kathleen K. King, "A Market Evaluation of the Risk-Based Capital Standards for the U.S. Financial System," *Journal of Banking and Finance*, 19: 531-562, 1995.

Coyne, Christopher, Ginette McManus, and A. J. Stagliano, "Data Envelopment Analysis and Management Performance: An Empirical Investigation of Small U.S. Banks," Small Firm Financial Research Symposium, California State University, Long Beach, 1993.

Demirguc-Kunt, Asli, "Deposit-Institution Failures: A Review of the Empirical Literature," Federal Reserve Bank of Cleveland, *Economic Review*, Quarter 4, 1989.

DeYoung, Robert, "Fee-Based Services and Cost Efficiency in Commercial Banks," Federal Reserve Bank of Chicago, *Proceedings: The Declining Role of Banks*, May: 501-519, 1994.

DeYoung, Robert, "X-Efficiency and Management Quality in National Banks," forthcoming, *Journal of Financial Services Research*.

DeYoung, Robert, and Daniel E. Nolle, "Foreign-Owned Banks in the U.S.: Earning Market Share or Buying It?," Office of the Comptroller of the Currency, Working Paper 95-2, March 1995.

DeYoung, Robert, and Gary Whalen, "Is a Consolidated Banking Industry a More Efficient Banking Industry?," Office of the Comptroller of the Currency, *Quarterly Journal*, volume 13, number 3, September 1994.

Eisenbeis, Robert, and Simon Kwan, "An Analysis of Inefficiencies in Banking: A Stochastic Cost Frontier Approach," working paper, 1994.

Elbadawi, I., A.R. Gallant, and G. Souza, "An Elasticity Can Be Estimated Consistently Without A Priori Knowledge of the Functional Form," *Econometrica*, 51: 1731-1753, 1983.

Federal Financial Institutions Examination Council, *Reports of Condition and Income*, 1984-1993.

Faulhaber, Gerald R., "Banking Markets: Productivity, Risk, and Customer Satisfaction," The Wharton School, University of Pennsylvania, working paper, 1995.

General Accounting Office, *Report to the Congress on the Bank Insurance Fund*, GAO/AFMD-90-100 (September 1990).

Hughes, Joseph P., William Lang, Loretta J. Mester, and Choon-Geol Moon, "Recovering Technologies that Account for Generalized Managerial Preferences: An Application to Non-Risk-Neutral Banks," forthcoming in *Proceedings: The New Tool Set, Assessing Innovations in Banking*, Federal Reserve Bank of Chicago, May 1995.

Hughes, Joseph P., and Loretta J. Mester, "A Quality and Risk-Adjusted Cost Function for Banks: Evidence on the 'Too-Big-To-Fail' Doctrine," *The Journal of Productivity Analysis*, 4: 293-315, 1993.

Hughes, Joseph P., and Choon-Geol Moon, "Measuring Bank Efficiency when Managers Trade Return for Reduced Risk," working paper, 1995.

Humphrey, David, and Lawrence Pulley, "Banks' Responses to Deregulation: Profits, Technology, and Efficiency," working paper, March 1995.

Hunter, William C., and Stephen Timme. "Core Deposits and Physical Capital: A Reexamination of Bank Scale Economies and Efficiency with Quasi-Fixed Inputs," *Journal of Money, Credit, and Banking*, 27: 165-185, 1995.

Jondrow, James, C.A. Knox Lovell, Ivan Materov, and Peter Schmidt, "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model," *Journal of Econometrics,* 19: 233-238, 1982.

Jones, David S., and Kathleen K. King, "The Implementation of Prompt Corrective Action," *Journal of Banking and Finance*, 19: 491-510, 1995.

McAllister, Patrick H., and Douglas McManus, "Resolving the Scale Efficiency Puzzle in Banking," *Journal of Banking and Finance*, 17: 389-405, 1993.

Mitchell, Karlyn, and Nur M. Onvural, "Economies of Scale and Scope at Large Commercial Banks: Evidence from the Fourier Flexible Functional Form," North Carolina State University, working paper, November 1992.

Mester, Loretta, "Efficiency of Banks in the Third Federal Reserve District," Federal Reserve Bank of Philadelphia, Working Paper No. 94-1, 1994a.

Mester, Loretta, "How Efficient Are Third District Banks?," Federal Reserve Bank of Philadelphia, *Business Review*, January 1994b.

Peristiani, Stavros, "Evaluating the Postmerger X-Efficiency and Scale Efficiency of U.S. Banks," Federal Reserve Bank of New York, working paper, 1993.

Resti, Andrea, "Linear Programming and Econometric Methods for Bank Efficiency Evaluation: An Empirical Comparison Based on a Panel of Italian Banks," working paper, 1995.

Stevenson, Rodney E., "Likelihood Functions for Generalized Stochastic Frontier Estimation," *Journal of Econometrics*, 13: 58-66, 1980.

Wall Street Journal, "Bank of Boston Top Managers Knew of Weak Loans Ahead of '89 Disclosure," April 25, 1994.

Whalen, Gary, "A Proportional Hazards Model of Bank Failure: An Examination of Its Usefulness as an Early Warning Tool," Federal Reserve Bank of Cleveland, *Economic Review*, Quarter 1, 1991.

Wheelock, David C. and Paul W. Wilson, "Explaining Bank Failures: Deposit Insurance, Regulation, and Efficiency," *Review of Economics and Statistics*, 77, 1995.

Appendix A

Intertemporal Patterns Underlying the Four Hypotheses

A. Bad Luck

1. A bank experiences bad luck (e.g., regional recession). Nonperforming loans increase as borrowers fall behind in their loan repayments.

2. After the increase in nonperforming loans, measured cost efficiency decreases because of extra costs of monitoring, workouts, default, etc.

3. Financial capital erodes in response to lost revenues, higher loan loss provisions, and greater costs associated with nonperforming loans.

Key Identifying Empirical Implication: Nonperforming loans *negatively* Granger-cause measured cost efficiency.

B. Bad Management

1. Managers: (a) fail to control costs, yielding low cost efficiency in the short term, and (b) are perform poorly at loan underwriting and monitoring.

2. Because of bad underwriting and monitoring, nonperforming loans eventually increase as unqualified and/or poorly monitored borrowers begin to fall behind in loan payments.

3. Equity capital erodes because cost inefficiency, lost revenues, and high loan loss provisions that depress earnings.

Key Identifying Empirical Implication: Measured cost efficiency *negatively* Granger-causes nonperforming loans.

Appendix A

C. Skimping

1. Managers choose to skimp on loan underwriting and monitoring, which reduces operating costs and increases measured cost efficiency in the short run.

2. Nonperforming loans increase in the long-run as poorly monitored borrowers fall behind in loan repayments.

3. Costs may increase and measured short-run cost efficiency may decrease due to extra costs of managing nonperforming loans.

4. Capital ratio may erode in response to the revenue loss, higher provisions, and possibly higher costs from nonperforming loans.

Key Identifying Empirical Implication: Measured short-run cost efficiency *positively* Granger-causes nonperforming loans.

D. Moral Hazard

1. Decline in earnings from any source -- including the first three hypotheses -- erodes financial capital.

2. In response to the low level of capital, the bank responds to moral hazard incentives by increasing loan risk.

3. As a result of the increase in risk-taking, banks eventually have higher nonperforming loans on average.

Key Identifying Empirical Implication: Capital ratio *negatively* Granger-causes nonperforming loans ratio.

Appendix B

Public Policy and Research Implications of the Four Hypotheses

A. Bad Luck

1. Regulation and supervision should continue to emphasize low portfolio credit risk (e.g., loan-to-asset ratios, risk-weighted asset ratios, loan concentration ratios, geographic diversification).

2. Bank failure research should focus on regional economic conditions as well as other traditional factors.

3. When estimating efficiency, the cost function, profit function, or production relationship should control for loan quality.

B. Bad Management

1. Bank supervision and bank failure research should include cost efficiency along with loan losses, credit risk, and other predictors of troubled banks.

2. When estimating efficiency, the cost function, profit function, or production relationship should *not* control for loan quality.

C. Skimping

1. Bank supervision and bank failure research should continue to concentrate on banks' credit control practices (e.g., loan review, collateral appraisal) as well as emphasizing other traditional factors.

2. Cost efficiency should be estimated only using long-run models and several years of data, to let the fluctuations caused by long-run cost strategies average out over time. Profit efficiency models may be superior to cost efficiency models.

3. When estimating efficiency, the cost function, profit function, or production relationship should *not* control for loan quality.

D. Moral Hazard

1. Bank supervision, bank regulation, and bank failure research should continue their traditional focus on the effects of capital, measuring capital, and setting and enforcing capital requirements.