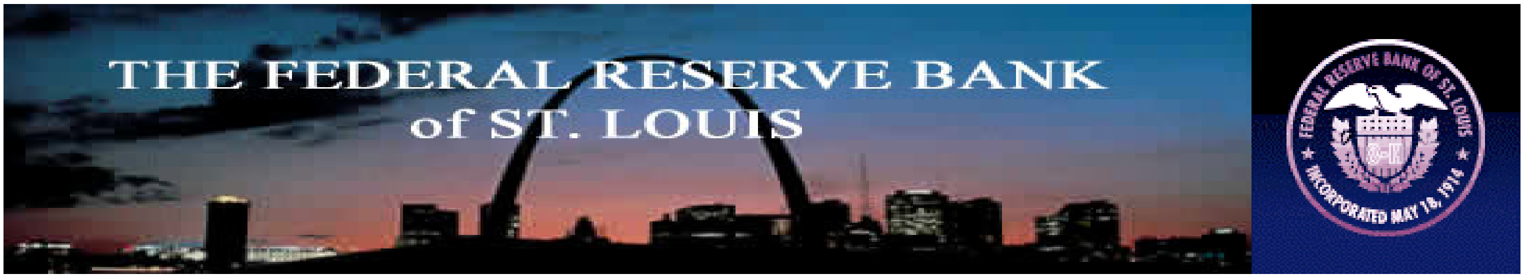


WORKING PAPER SERIES

The Role of a CAMEL Downgrade Model in Bank Surveillance

R. Alton Gilbert, Andrew P. Meyer and Mark D.Vaughan

Working Paper 2000-021A
http://research.stlouisfed.org/wp/2000/2000-021.pdf

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
411 Locust Street
St. Louis, MO 63102

THE ROLE OF A CAMEL DOWNGRADE MODEL IN BANK SURVEILLANCE

Abbreviated title: The Role of a CAMEL Downgrade Model in Bank Surveillance

R. Alton Gilbert
Research Department
Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166
314-444-8559
FAX 314-444-8731
gilbert@stls.frb.org

Andrew P. Meyer
Banking Supervision and Regulation
Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166
314-444-4647
ameyer@stls.frb.org

Mark D. Vaughan
Banking Supervision and Regulation
Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166
314-444-8859
mvaughan@stls.frb.org

ABSTRACT

This article examines the potential contribution to bank supervision of a model designed to predict which banks will have their supervisory ratings downgraded in future periods. Bank supervisors rely on various tools of off-site surveillance to track the condition of banks under their jurisdiction between on-site examinations, including econometric models. One of the models that the Federal Reserve System uses for surveillance was estimated to predict bank failures. Because bank failures have been so rare during the last decade, the coefficients on this model have been "frozen" since 1991. Each quarter the surveillance staff at the Board of Governors provide the supervision staff in the Reserve Banks the probabilities of failure by the banks subject to Fed supervision, based on the coefficients of this bank failure model and the latest call report data for each bank. The number of banks downgraded to problem status in recent years has been substantially larger than the number of bank failures. During a period of few bank failures, the relevance of this bank failure model for surveillance depends to some extent on the accuracy of the model in predicting which banks will have their supervisory ratings downgraded to problem status in future periods. This paper compares the ability of two models to predict downgrades of supervisory ratings to problem status: the Board staff model, which was estimated to predict bank failures, and a model estimated to predict downgrades of supervisory ratings. We find that both models do about as well in predicting downgrades of supervisory ratings for the early 1990s. Over time, however, the ability of the downgrade model to predict downgrades improves relative to that of the model estimated to predict failures. This pattern reflects the value of using a model for surveillance that can be re-estimated frequently. We conclude that the downgrade model may prove to be a useful supplement to the Board's model for estimating failures during periods when most banks are healthy, but that the downgrade model should not be considered a replacement for the current surveillance framework.

Key words: bank supervision, early warning models, and off-site surveillance.

JEL Codes: G21, G28, and C53

I. INTRODUCTION

Banking is one of the more closely supervised industries in the United States, reflecting the view that bank failures have stronger adverse effects on economic activity than other business failures. The federal government and the state governments grant authority to bank supervisors to limit the risk of failure assumed by banks. Supervisors impose sanctions on the banks that they have identified as being in poor financial condition. Effective bank supervision, therefore, requires accurate information about the condition of banks.

Bank supervisors use on-site examination and off-site surveillance to identify the banks most likely to fail. The most useful tool for identifying problem institutions is on-site examination, in which examiners travel to a bank and review all aspects of its safety and soundness. On-site examination is, however, both costly and burdensome: costly to supervisors because of its labor-intensive nature and burdensome to bankers because of the intrusion into their day-to-day operations. As a result, supervisors also monitor bank condition off-site. Off-site surveillance yields an ongoing picture of bank condition, enabling supervisors to schedule and plan exams efficiently. Off-site surveillance also provides banks with incentives to maintain safety and soundness between on-site visits.

Supervisors rely primarily on two analytical tools for off-site surveillance: supervisory screens and econometric models. Supervisory screens are combinations of financial ratios, derived from bank balance sheets and income statements, that have, in the past, given forewarning of safety-and-soundness problems. Supervisors draw on their experience to weigh the information content of these ratios.

One of the contributions of economists to bank supervision has been the estimation and simulation of econometric models designed to provide supervisors with early warning of the banks that are most likely to develop serious problems in the future. Econometric models use the information about the condition of banks in their financial statements to derive one number. In most early warning models, that number is the probability that a bank will fail in a future period. In a recent article the authors compared the accuracy of supervisory screens and econometric models in predicting which banks would fail in future periods, and in predicting which banks would have their supervisory ratings downgraded to problem bank status (Gilbert, Meyer and Vaughan, 1999). Our analysis and other research demonstrate that econometric models dominate supervisory screens as predictors of bank failure and downgrades of supervisory ratings.

The Federal Reserve System uses the System to Estimate Examination Ratings, or SEER model, as one of its principal off-site surveillance tools. Surveillance staff at the Board of Governors use one form of this model, the risk rank model, to compute a probability of failure for each bank. Because bank failures have been so rare during the last decade, the coefficients on this model have been "frozen" since 1991. Each quarter supervisors at each of the twelve Reserve Banks receive information from the surveillance staff of the Board of Governors about the probabilities of failure by the banks that are supervised by Federal Reserve staff. Probabilities of failure are calculated using the latest call report data for each bank and the "frozen" coefficients of the model estimated to predict bank failure.

This paper investigates a practical issue in the use of econometric models in bank surveillance: the number of models that are relevant for surveillance. That is, does a model estimated to predict which banks are most likely to fail in future periods also provide accurate predictions of which banks are most likely to be downgraded to problem bank status? Or do supervisors need a different model to predict which banks they are most likely to rate as problem banks in future periods? Ability to predict downgrades of supervisory ratings, rather than failures, is especially relevant for surveillance during a period in which there are few failures, like most of the 1990s. We compare the performance of two econometric models in predicting which banks will have their supervisory ratings downgraded in future periods: the SEER risk rank model, which was estimated to predict bank failures, and another model that is estimated to predict downgrades of supervisory ratings.

II. ON-SITE AND OFF-SITE SURVEILLANCE: A CLOSER LOOK

This section discusses the role of off-site surveillance and early warning models in bank supervision. Bank supervisors rely principally on regular on-site examinations to assess the condition of banks. Examinations ensure the integrity of bank financial statements and identify the banks that should be subject to supervisory sanctions. During a routine exam, the examiners assess six components of safety and soundness—capital protection (C), asset quality (A), management competence (M), earnings strength (E), liquidity risk (L) and market risk (S)—and assign a grade of 1 (best) through 5 (worst) to each component. Examiners then use these six scores to award a composite rating, also expressed on a 1 through 5 scale. Bank supervisors added the "S" component (market risk) in January 1997. Since examiners graded only five components of safety and soundness during most of our sample period, this paper refers to composite "CAMEL" ratings. Table 1 interprets the five composite CAMEL ratings.

TABLE 1 ABOUT HERE

Although on-site examination is the most effective tool for constraining bank risk, it is both costly to supervisors and burdensome to bankers. As a result, supervisors face continuous pressure to limit exam frequency. Supervisors yielded to this pressure in the 1980s, and many banks escaped yearly examination (Reidhill and O'Keefe, 1997). Congress mandated the frequency of examinations in the Federal Deposit Insurance Corporation Improvement Act of 1991, which requires annual examinations for all but a handful of small, well-capitalized, highly-rated banks, and even these institutions must be examined every 18 months. This new mandate reflects the lessons learned from the wave of bank failures in the late 1980s: more frequent exams, though likely to increase the up-front costs of supervision, reduce the down-the-road costs of resolving failures by revealing problems at an early stage.

Although changes in public policy have mandated greater exam frequency since the early 1990s, supervisors still have reasons to use off-site surveillance tools to flag banks for accelerated exams and to plan exams. Bank condition can deteriorate rapidly between on-site visits (Cole and Gunther, 1998**;** Hirtle and Lopez, 1999). In addition, the Federal Reserve now employs a "risk-focused" approach to exams, in which supervisors allocate on-site resources according to the risk exposures of each bank (Board of Governors, 1996). Off-site surveillance helps supervisors allocate on-site resources efficiently by identifying institutions that need immediate attention and by identifying specific risk exposures for regularly scheduled as well as accelerated exams. For these reasons, an interagency body of bank and thrift supervisors—the Federal Financial Institutions Examinations Council (FFIEC)—requires banks to submit quarterly Reports of Condition and Income, often referred to as the call reports. Surveillance analysts use the call report data to monitor the condition of banks between exams.

Supervisors have developed various tools for using call report data to schedule and plan exams, including econometric models. A common type of model used in surveillance estimates the marginal impact of a change in a financial ratio on the probability that a bank will fail, holding all other ratios constant. These models can examine many ratios simultaneously, capturing subtle but important interactions. The Federal Reserve uses two models in off-site surveillance. One model, called the SEER risk rank model, combines financial ratios to estimate the probability that each Fed-supervised bank will fail within the next two years. Another model estimates a hypothetical CAMEL rating that is consistent with the financial data in the bank's most recent call report. Every quarter, economists at the Board of Governors feed the latest call report data into these models and forward the results to each of the twelve Reserve Banks. Surveillance analysts in the Reserve Banks then investigate the institutions that the models flag as "exceptions."

III. ESTIMATION OF THE SEER RISK RANK MODEL

The SEER risk rank model is a probit model of bank failure. The model uses call report data as of one point in time to estimate the probability of banks failing or becoming critically undercapitalized over the following two years. The estimation period spans the first quarter of 1985 to the last quarter of 1991. Table 2 lists the independent variables used in this failure prediction model. The actual SEER coefficient estimates, which are part of the official Federal Reserve monitoring system, are not reported here because they are confidential.

TABLE 2 ABOUT HERE

Because bank failures have been rare events since the early 1990s, the regression coefficients have been “frozen” since 1991, rather than being re-estimated in more recent years. Simulation of this model is relevant for bank surveillance during a period of few bank failures if the determinants of bank failure are also the determinants of CAMEL downgrades, and if the weights on the various determinants of distress in banks included in the model do not change much over time. This paper investigates whether the ranking of banks by their estimated probabilities of failure (from the SEER risk rank model) provides reliable predictions of the banks most likely to have their CAMEL ratings downgraded in future periods.

IV. ESTIMATION OF THE DOWNGRADE MODEL

To facilitate a direct comparison between the performance of the two models in predicting CAMEL downgrades, we used the same independent variables in estimation of the SEER risk rank model and the downgrade equations, which are listed in Table 2. If we used different independent variables in estimating the two models, we would not know whether differences in the performance of the two models reflected different independent variables or the estimation of the models with different dependent variables (dummy variables for bank failures versus CAMEL downgrades).

We estimated the downgrade model for six separate years using probit regression analysis. Because there were considerably more CAMEL downgrades than failures in the

1990s, it was possible to re-estimate the downgrade model on a yearly basis. See Table 3 for the number of downgrades each year.

TABLE 3 ABOUT HERE

We chose the timing of CAMEL ratings and call report data for the estimation and simulation of the downgrade model to reflect the timing of the information that is available to supervisors in practice. Banks included in the sample for the first equation in Table 4 were rated CAMEL 1 or 2 as of March 1990. Call report data for the independent variables listed in Table 2 were as of the fourth quarter of 1989 because supervisors would not have access to the fourth quarter 1989 call report data for most banks until some time in March 1990. We assume that the supervisors would like to be able to rank these banks currently rated CAMEL 1 or 2 by their probability of being downgraded during the year 1991. Supervisors also would like to have accurate information on the banks most likely to be downgraded in the remainder of 1990, but early intervention would require accurate information on the banks most likely to be downgraded in 1991.

The dependent variable of the first equation had a value of unity if a bank had its CAMEL rating downgraded from a 1 or 2 (satisfactory condition) to a CAMEL 3, 4 or 5 (problem bank status) in the calendar year 1991, zero otherwise. We limited the sample to banks that were examined at least once during the calendar year 1991 because supervisors generally change the CAMEL ratings of banks only after exams. We also excluded banks from the sample if they were downgraded to problem bank status in April through December of 1990, since these banks would not be included among the banks downgraded during the year 1991. The other five equations in Table 4 involve banks that were downgraded in the calendar years 1992 through 1996, with the same timing of the observations on CAMEL ratings and call report data.

Using the same independent variables in the SEER risk rank model and this downgrade model tends to limit the predictive power of the downgrade model. It is likely that an attempt to estimate the optimal downgrade model would yield at least some independent variables not included in the SEER risk rank model. For instance, some of the independent variables with significant coefficients in the downgrade model in Gilbert, Meyer and Vaughan (1999) are not included as independent variables in the downgrade model in this paper. Thus, the downgrade model in this paper is not designed to minimize the errors in predicting which banks will have their CAMEL ratings downgraded. Instead, it is designed to facilitate a comparison of the SEER risk rank model to a downgrade model.

TABLE 4 ABOUT HERE

Results in Table 4 indicate that the model fit for the downgrade regressions is very good. The coefficients on seven of the 11 variables are statistically significant at the five-percent level in each of the six years, with the signs indicated in table 2. The fit of the model declines gradually each year, however, with more variables becoming statistically insignificant later in the 1990s. For instance, the coefficients on the ratio of commercial and industrial loans to total assets were negative and statistically significant at the five-percent level in the years 1991 through 1993, but insignificant in 1994 through 1996. The coefficients on equity to total assets were positive and statistically significant in the years 1991 through 1994, but insignificant in 1995 and 1996. The poorer fit of the downgrade model over time may reflect the decline in the number of downgrades throughout the 1990s.

V. OUT-OF-SAMPLE MODEL PERFORMANCE

While a good in-sample model fit is important for determining whether a model is statistically well specified, the real test for use in surveillance is its out-of-sample predictive performance. We use out-of-sample simulation to answer this question: Using a model estimated with existing information, how well can we distinguish the banks that will be downgraded in a future period from those that will not be downgraded?

The simulations are designed to reflect the information that supervisors could have used at the time for early warning of future CAMEL downgrades. The first out-of-sample simulation involved banks rated CAMEL 1 or 2 as of March 1992. We estimated the probabilities that these banks would be downgraded to problem bank status in 1993 by plugging their call report data as of the fourth quarter of 1991 into the equation estimated to predict downgrades in 1991. The other out-of-sample simulations of the downgrade model use the same lags for the other years. For instance, we used the coefficients on the sixth equation, estimated for banks downgraded in 1996, to estimate the probabilities that the banks rated CAMEL 1 or 2 as of March 1997 would be downgraded in 1998.

We used the following procedure to predict which banks would be downgraded to problem bank status with the SEER risk rank model. We ranked the banks in our sample by their probability of failing in 1993 by plugging their call report data as of the fourth quarter of 1991 into the SEER risk rank model and calculating the estimates of their probabilities of failing. We used this ranking by their probabilities of failing as a means of ranking these banks by their probabilities of being downgraded to problem bank status in 1993. To derive a ranking of banks by their probability of being downgraded in 1994, we plugged their call report data as of the fourth quarter of 1992 into the SEER risk rank model with its frozen coefficients. We followed this procedure for each year, using the model with the same coefficients and the new call report data.

In evaluating the out-of-sample performance of these models, we must consider the trade-off between type-1 and type-2 errors. The type-1 error rate is the percentage of downgrades that we fail to detect with the model, while the type-2 error rate is the percentage of healthy banks that we identify incorrectly as downgrade risks. In a traditional bank *failure* analysis, supervisors consider type-1 errors to be decidedly worse than type-2 errors. A bank failure imposes significant costs on society, including required payouts by the FDIC insurance fund (and possibly the taxpayers) and possible disruption to the bank's local community (Gilbert and Kochin, 1989). An early-warning system that misses a significant number of failures (i.e., has a high type-1 error rate) is highly undesirable. Errors in predicting CAMEL downgrades in future periods are considerably less serious than errors in predicting bank failures, but supervisors are interested in tools that provide accurate early warnings of CAMEL downgrades, which could enable them to provide added guidance to problem banks as early as possible.

On the other hand, a model that over-predicts failures and downgrades (i.e., has a high type-2 error rate) also imposes costs in the form of unnecessary resources devoted to the supervision of healthy banks. These costs range from a few extra examiner hours spent examining bank financial statements to the cost of stepped-up on-site examinations. Type-2 errors also **may** impose unnecessary costs on healthy banks. An explicit calculation of the costs of each type of error, which would involve certain arbitrary assumptions, is beyond the scope of this paper. We choose instead to focus on the "risk-return" trade-off along the entire range of errors by deriving type-1 vs. type-2 power curves for the two models, following closely the methods used by Cole, Cornyn and Gunther (1995). Each curve tells us the type-1 error we must accept for any given level of type-2 error. Looking at the entire power curve avoids setting an arbitrary cut-off for an acceptable level of type-1 or type-2 error.

To compute power curves for the SEER risk rank model and the downgrade model, we adjust the number of banks rated by each model as likely to be downgraded and observe the associated type-1 and type-2 errors. At one extreme, all banks are rated as unlikely to be downgraded. Our type-1 error rate would be 100 percent, since all of the downgrades would be recorded as errors. The type-2 error rate would be zero, since each bank that was not downgraded would be included among those not predicted to be downgraded. Next, we assume that the bank with the highest estimated probability of failing (from the SEER risk rank model) or being downgraded (from the downgrade model) is the one bank predicted to be downgraded in the future period. For each model we calculate the associated type-1 and type-2 errors. We continue adding the bank with the next highest estimated probability of failure or being downgrade to those already estimated to be downgraded, and calculating type-1 and type-2 errors, until we have traced out the entire power curve of type-1 and type-2 errors. At the other extreme position on the power curve, all banks are included among those estimated to be downgraded in the future; the type-1 error rate is zero percent, and the type-2 error rate is 100 percent.

The curvature of the power curves provides a basis for comparing the performance of the two models. The greater the curvature of the power curve, the better the model. In other words, we can simultaneously achieve low type-1 and type-2 error rates using the model with the curve nearer the origin. A convenient way to express this curvature is to calculate the area under each power curve as a percentage of all of the area in the box where the power curves are presented. The smaller the area, the more accurate the model. A useful benchmark is the case in which the banks estimated to be downgraded in the future are selected at random, rather than through simulation of a model. This procedure would produce a power curve with a slope of approximately negative one, starting at the 100 percent type-1 error rate and extending to the 100 percent type-2 error rate. The area under this curve would be approximately 50 percent of the area in the entire box. The areas under power curves in Figures 1 through 6 represent a clear improvement over a random selection of banks estimated to be downgraded.

FIGURES 1 THROUGH 6 ABOUT HERE

The power curves in Figure 1 are virtually on top of each other, with the downgrade curve slightly edging out the failure curve over most of the trade-off range in 1993. The area under the power curve in Figure 1 for the downgrade model is 20.31 percent, only slightly better than the 21.47 percent for the SEER risk rank model. The similarity of these curves indicates the close relationship between the risk factors that cause failures and those that cause less significant deterioration in the condition of banks. Figure 2 shows a similar pattern for 1994 downgrades. The spread between the power curves starts to widen by 1995, however, and peaks at 5.12 percentage points in 1997.

Table 5 summarizes the information about the areas under the power curves for each year.

TABLE 5 ABOUT HERE

The increase in the relative performance of the downgrade model over time has several implications. First, this pattern could indicate that the coefficients that assign weights to the various risk factors are not stable over time. If some variables were major predictors of bank performance in some years, but not others, a model with frozen coefficients would decline in usefulness over time. Evidence of coefficient instability over long periods indicates the value of re-estimating the model from time to time. During the 1990s, re-estimation has been possible for the downgrade model, but not for the failure model. A second implication is that the relative predictive power of the downgrade model increased over time even as its in-sample statistical properties deteriorated. Re-estimation of the coefficients of the downgrade model over time enhanced the predictive power of the downgrade model, even with the deterioration over time of the in-sample statistical properties of the downgrade model. These results indicate that during long periods of economic expansion when there are few bank failures, a model estimated to predict the downgrades of CAMEL ratings can serve as a useful supplement to an early warning model estimated to predict bank failures.

VI. CONCLUSIONS

This paper demonstrates that the SEER risk rank model, which was estimated to predict which banks will fail, does a good job of identifying the banks that are likely to have their supervisory ratings downgraded in future periods. During a period with few bank failures, however, a model estimated to predict CAMEL downgrades could improve on the ability of the SEER risk rank model to predict downgrades. We must recognize, however, that our sample period is taken from the longest peacetime expansion in U.S. economic history. Consequently, a model designed to predict which healthy banks will become unhealthy may do a poor job of predicting which unhealthy banks will ultimately fail. While a downgrade model may prove to be a useful supplement to the standard failure prediction model during good times, we should be careful to have all of our surveillance tools available to us during the next recession.

Use of both a bank failure model and a CAMEL downgrade model in bank surveillance may be better than choosing to use one of these models. Running more than one model in parallel could help us examine more closely the causes of bank downgrades that may be quite different from the causes of outright failure. Furthermore, while a downgrade model might serve admirably during periods of relative stability in the banking system, a failure model might be a more important tool for surveillance during a future period when the bank failure rate is higher than in recent years.

ACKNOWLEDGMENT

The authors thank Robert Avery, Kevin Bertsch, Donald Conner, William Francis, Michael Gordy, Jeffrey Gunther, James Harvey, Stephen Jenkins, Kim Nelson and Joseph Turk for helpful comments. Boyd Anderson and Thomas King provided excellent research assistance. The views expressed in this paper do not necessarily represent those of the Federal Reserve Bank of St. Louis, the Board of Governors, or the Federal Reserve System.

NOTES

1. See Flannery and Houston (1999) for evidence that holding company inspections help insure the integrity of financial statements. See Gilbert and Vaughan (1998) for a discussion of the sanctions available to bank supervisors.

2. See Hall, King, Meyer and Vaughan (1999) for a more detailed discussion of the factors used to assign individual and composite ratings.

3. See Putnam (1983) for a description of the use of supervisory screens in off-site surveillance during the late 1970s and early 1980s.

REFERENCES

Board of Governors of the Federal Reserve System (1996). Risk-Focused Safety and Soundness Examinations and Inspections. SR 96-14, May 24.

Cole, R.A., Cornyn, B.G. & Gunther, J.W. (1995). FIMS: A New Monitoring System for Banking Institutions. Federal Reserve Bulletin, 81, 1-15.

Cole, R.A. & Gunther, J.W. (1998). A Comparison of On- and Off-Site Monitoring Systems. Journal of Financial Services Research, 13, 103-17.

Flannery, M.J. & Houston, J.F. (1999). The Value of a Government Monitor for U.S. Banking Firms. Journal of Money, Credit, and Banking, 31, 14-34.

Gilbert, R.A., & Kochin, L.A. (1989). Local Economic Effects of Bank Failures. Journal of Financial Services Research, 3, 333-45.

Gilbert, R.A., Meyer, A.P. & Vaughan, M.D. (1999). The Role of Supervisory Screens and Econometric Models in Off-Site Surveillance. Federal Reserve Bank of St. Louis Review, 81, 31-56.

Gilbert, R.A. & Vaughan, M.D. (1998). Does the Publication of Supervisory Enforcement Actions Add to Market Discipline? Research in Financial Services: Public and Private Policy, 10, 259-80.

Hall, J.R., King, T.B., Meyer, A.P. & Vaughan, M.D. (1999). Do Certificate of Deposit Holders and Supervisors View Bank Risk Similarly? A Comparison of the Factors Affecting CD Yields and CAMEL Composites. Supervisory Policy Analysis Working Paper, Federal Reserve Bank of St. Louis, October.

Hirtle, B.J. & Lopez, J.A. (1999). Supervisory Information and the Frequency of Bank Examinations. Economic Policy Review, Federal Reserve Bank of New York, 5, No. 1, 1-19.

Putnam, B.H (1983). Early Warning Systems and Financial Analysis in Bank Monitoring: Concepts of Financial Monitoring. Federal Reserve Bank of Atlanta Economic Review, November, 6-13.

Reidhill, J. & O'Keefe, J.O. (1997). Off-Site Surveillance Systems, in History of the Eighties: Lessons for the Future, volume I. Washington: Federal Deposit Insurance Corporation, 477-520.

Table 1. Interpretation of CAMEL Composite Ratings

CAMEL Composite Rating	Description
1	Financial institutions with a composite 1 rating are sound in every respect and generally have individual component ratings of 1 or 2.
2	Financial institutions with a composite 2 rating are fundamentally sound. In general, a 2-rated institution will have no individual component ratings weaker than 3.
3	Financial institutions with a composite 3 rating exhibit some degree of supervisory concern in one or more of the component areas.
4	Financial institutions with a composite 4 rating generally exhibit unsafe and unsound practices or conditions. They have serious financial or managerial deficiencies that result in unsatisfactory performance.
5	Financial institutions with a composite 5 rating generally exhibit extremely unsafe and unsound practices or conditions. Institutions in this group pose a significant risk the deposit insurance fund and their failure is highly probable.

Source: *Federal Reserve Commercial Bank Examination Manual*

Table 2. Regression Variables Used to Predict Bank Failures and CAMEL Downgrades

This table lists the independent variables used in both the SEER risk rank model and the downgrade regression model. The signs indicate the hypothesized relationships between the variables and the likelihood of a safety-and-soundness problem. For example, the negative sign for the return-on-assets ratio indicates that other things equal, a higher ROA would reduce the likelihood of a failure or CAMEL downgrade.

Symbol	Description	Hypothesis about the sign of the coefficient for predicting failure or CAMEL downgrades (positive sign indicates positive correlation with probability of failure or rating downgrade).
ROA	Return on average assets.	–
COMMERCIAL-LOANS	Commercial and industrial loans as a percentage of total assets.	+
NET-WORTH	Total net worth (equity capital minus goodwill) as a percentage of total assets.	–
OREO	Other real estate owned as a percentage of total assets.	+
PAST-DUE-30	Loans past due 30-89 days as a percentage of total assets.	+
PAST-DUE-90	Loans past due 90+ days as a percentage of total assets.	+
NONACCRUING	Nonaccrual loans as a percentage of total assets.	+
SECURITIES	Book value of securities as a percentage of total assets.	–
LARGE-TIME-DEPOSITS	Deposits > $100M (jumbo CDs) as a percentage of total assets.	+
RESIDENTIAL-LOANS	Residential real estate loans as a percentage of total assets.	–
SIZE	Natural logarithm of total assets, in thousands of dollars.	–

Table 3. Banks Rated CAMEL 1 or 2 that were Downgraded to CAMEL 3, 4, or 5

This table shows the number of our sample banks that were downgraded from CAMEL 1 or 2 to 3, 4 or 5 in each year. We excluded from the sample any banks that received downgrades to problem status the same year as the CAMEL 1 or 2 observation. As overall banking performance improved in the mid-1990s, the percentage of banks suffering downgrades fell, but there was an upward trend in the late 1990s, and downgrades were still much more common than failures.

Date of Rating (Year of Downgrade)	CAMEL Rating	Number of Banks	Number of Banks Downgraded	Percentage Downgraded
March 1990 (1991)	1	2,057	79	3.84
	2	5,037	988	19.60
March 1991 (1992)	1	1,956	51	2.61
	2	4,987	670	13.43
March 1992 (1993)	1	1,978	17	0.86
	2	5,246	295	5.62
March 1993 (1994)	1	2,046	14	0.68
	2	5,040	185	3.67
March 1994 (1995)	1	2,363	13	0.55
	2	4,463	129	2.89
March 1995 (1996)	1	2,596	13	0.50
	2	3,964	136	3.43
March 1996 (1997)	1	2,655	13	0.49
	2	3,394	157	4.63
March 1997 (1998)	1	2,574	27	1.05
	2	2,803	194	6.92

Table 4. In-Sample Fit of the Downgrade Model

This table presents the estimated regression coefficients for the downgrade model. The model predicts in-sample downgrades ("1" represents a downgrade; "0" denotes no downgrade) for calendar year *t* with call report data for the fourth quarter of year *t-2*. For example, observations on whether banks had their CAMEL ratings downgraded in 1993 were related to call report data for the fourth quarter of 1991. Standard errors appear in parentheses below the coefficients. Three asterisks denote significance at the one-percent level; two asterisks denote significance at the 5-percent level. Shading highlights coefficients that were significant with the correct sign in all six years. Overall, the evidence in this table suggests that the logit model predicted in-sample downgrades well.

	Years of downgrades in CAMEL ratings:		
Independent Variables	**1991**	**1992**	**1993**
Intercept	-1.452***	0.700	0.599
	(0.435)	(0.506)	(0.705)
COMMERCIAL-LOANS	0.018***	0.029***	0.025***
	(0.005)	(0.006)	(0.008)
RESIDENTIAL-LOANS	-0.006	-0.001	-0.004
	(0.004)	(0.005)	(0.007)
LARGE-TIME-DEPOSITS	0.035***	0.037***	0.041***
	(0.005)	(0.006)	(0.009)
NET-WORTH	-0.108***	-0.090***	-0.087***
	(0.020)	(0.022)	(0.029)
PAST-DUE-90	0.712***	0.583***	0.614***
	(0.075)	(0.078)	(0.101)
PAST-DUE-30	0.243***	0.281***	0.238***
	(0.041)	(0.043)	(0.058)
NONACCRUING	0.439***	0.255***	0.702***
	(0.056)	(0.065)	(0.079)
ROA	-0.529***	-0.598***	-0.375***
	(0.073)	(0.082)	(0.098)
SECURITIES	-0.040***	-0.045***	-0.037***
	(0.004)	(0.004)	(0.006)
OREO	0.300***	0.273***	0.241***
	(0.051)	(0.057)	(0.063)
SIZE	0.074**	-0.171***	-0.291***
	(0.030)	(0.038)	(0.057)
Number of Observations	7,121	6,975	7,257
-2 log likelihood testing whether all coefficients (except the intercept) = 0	4824.234***	3700.348***	2073.600***

COMMERCIAL-LOANS	Commercial and industrial loans as a percentage of total assets	PAST-DUE-30	Loans over 30 days past due as a percentage of total loans
RESIDENTIAL-LOANS	Residential real-estate loans as a percentage of total assets	NONACCRUING	Loans on nonaccrual status as a percentage of total loans
LARGE-TIME-DEPOSITS	Large denomination time deposit liabilities as a percentage of total assets.	ROA	Net income as a percentage of total assets.
		SECURITIES	Book value of securities as a percentage of total assets
NET-WORTH	Equity less goodwill as a percentage of total assets	OREO	Other real estate owned as a percentage of total assets
PAST-DUE-90	Loans over 90 days past due as a percentage of total loans	SIZE	Natural logarithm of total assets, in thousands of dollars.

Table 4 (continued). How Well Does the Downgrade Model Fit the CAMEL Downgrade Data?

	Years of downgrades in CAMEL ratings:		
Independent Variables	**1994**	**1995**	**1996**
Intercept	1.589 (0.865)	0.611 (0.991)	3.160)*** (1.116)
COMMERCIAL-LOANS	0.004 (0.011)	0.009 (0.013)	0.015 (0.014)
RESIDENTIAL-LOANS	-0.013 (0.008)	-0.003 (0.009)	-0.037)*** (0.011)
LARGE-TIME-DEPOSITS	0.043*** (0.011)	0.063*** (0.012)	0.049)*** (0.013)
NET-WORTH	-0.107*** (0.032)	-0.037 (0.034)	-0.037 (0.037)
PAST-DUE-90	0.477*** (0.119)	0.396*** (0.147)	0.634)*** (0.208)
PAST-DUE-30	0.317*** (0.082)	0.261*** (0.080)	0.254)*** (0.091)
NONACCRUING	0.322*** (0.104)	0.468*** (0.114)	0.482)*** (0.127)
ROA	-0.521*** (0.112)	-0.509*** (0.126)	-0.329** (0.154)***
SECURITIES	-0.016** (0.006)	-0.019*** (0.008)	-0.042)*** (0.008)
OREO	0.304*** (0.096)	0.345*** (0.098)	0.333)*** (0.143)
SIZE	-0.362*** (0.073)	-0.380*** (0.087)	-0.532)*** (0.095)
Number of Observations	7,121	6,853	5,935
-2 log likelihood testing whether all coefficients (except the intercept) = 0	1654.108***	1241.822***	1008.453***

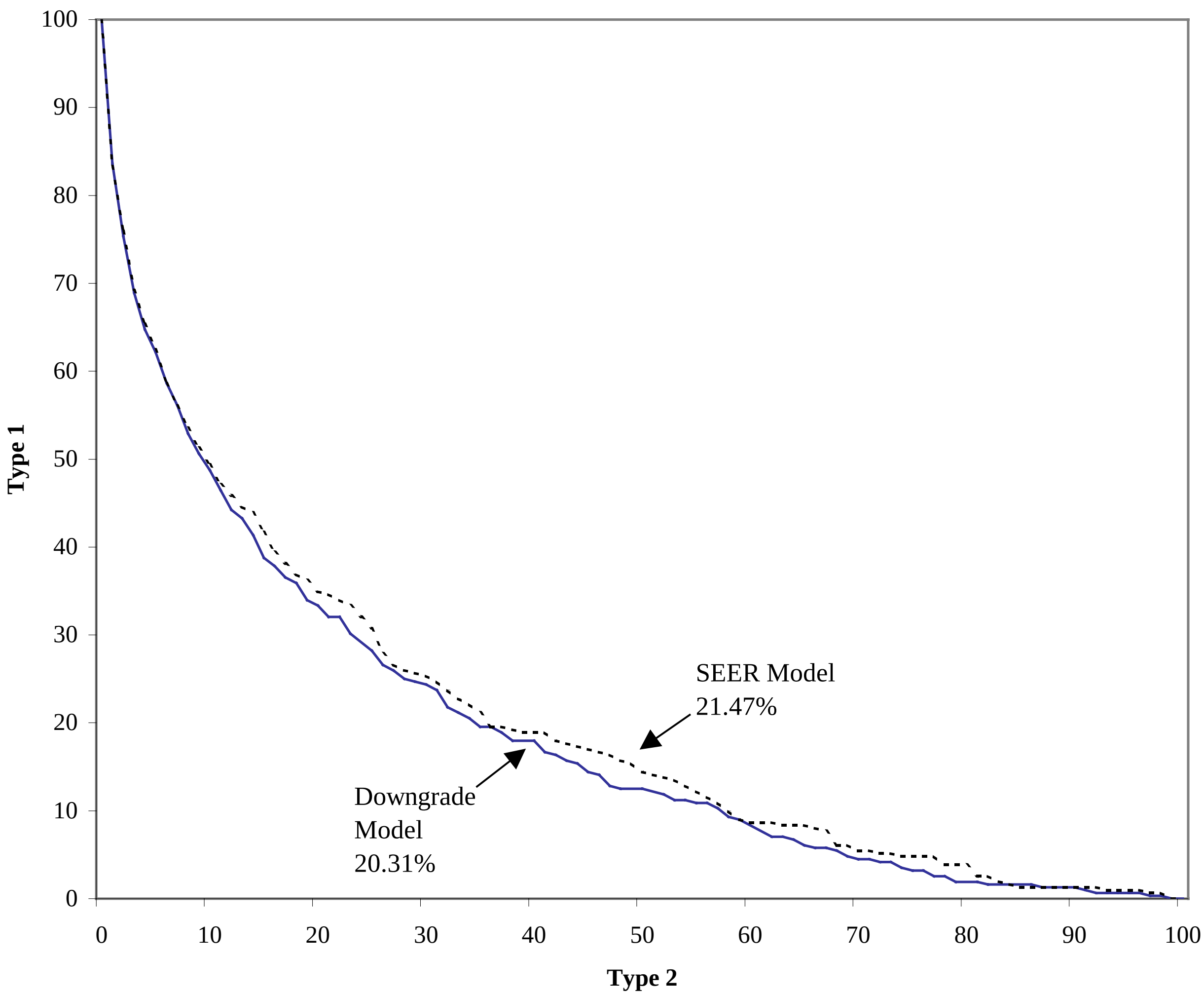
Table 5. Out-of-Sample Predictive Performance of the SEER Risk Rank Model and the Downgrade Model

This table shows the areas under the "power curves" in Figures 1 through 6, along with the percentage-point differences between the areas for the two models. A power curve graphically illustrates the trade-off between type-1 errors (the percentage of missed downgrades) and type-2 errors (the percentage of misidentified non-downgrades). A convenient way to compare curves is to calculate the area under each curve. In this context, smaller is better because smaller areas imply simultaneously lower error rates of both types. As shown below, the difference in the areas increases slowly over time, indicating that the downgrade model's performance, relative to the SEER risk rank model, improves as the SEER coefficients become increasingly "stale."

Figure Number	Downgrade Year	Downgrade Model Area	SEER Failure Model Area	Percentage-point Difference
1	1993	20.31%	21.47%	1.16%
2	1994	29.49%	30.66%	1.17%
3	1995	24.44%	26.87%	2.43%
4	1996	20.07%	24.31%	4.24%
5	1997	19.91%	25.03%	5.12%
6	1998	23.02%	25.81%	2.79%

Figure 1: What is the Trade-off Between False Negatives and False Positives in the Downgrade Prediction Model Compared to the SEER Model?

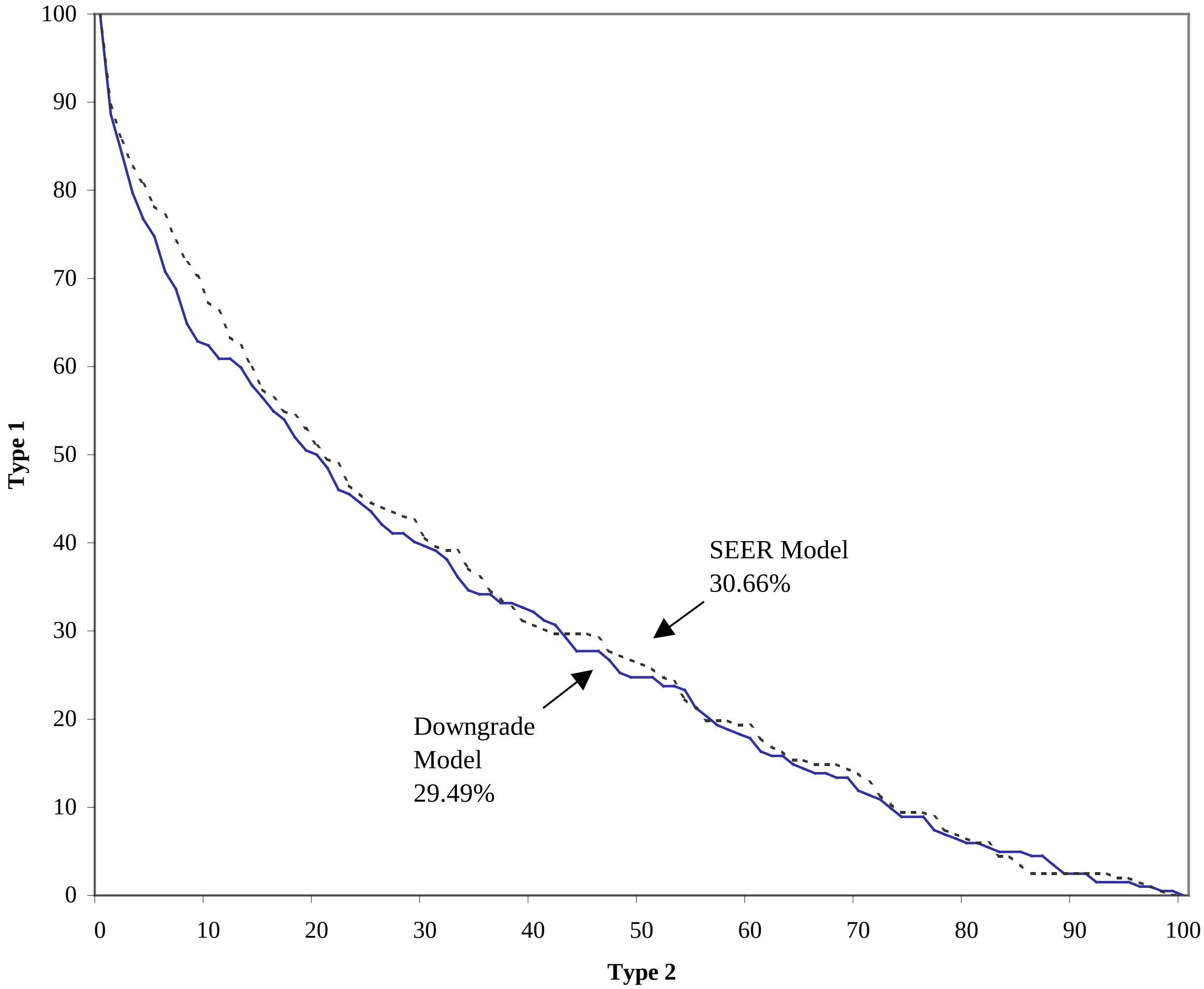
1993 Downgrade Predictions Using Year-end 1991 Data



This figure shows the trade-off between the type-1 error rate and the type-2 error rate. The type-1 error rate is the percentage of banks rated CAMEL-1 or -2 that were subsequently downgraded by supervisors but were not identified by the given model. The type-2 error rate is the percentage of banks rated CAMEL-1 or -2 that were not subsequently downgraded but were misidentified by the given model as a downgrade risk. A desirable early-warning system minimizes the level of type-2 errors for any given level of type-1 errors. This graph shows that the SEER risk rank model and the downgrade model have about the same type-1 and type-2 tradeoffs for any level of the type-1 error above 50 percent. For most levels of type-1 error below 50 percent, the downgrade model yields a lower type-2 error rate than the SEER risk rank model. A convenient way to express this difference in performance is to calculate the percentage of the area in the figure that is under each curve. Smaller areas are more desirable because they imply simultaneously low levels of both types of errors. In the above graph, the downgrade model slightly edges out the SEER model by 20.31 percent to 21.47 percent.

Figure 2: What is the Trade-off Between False Negatives and False Positives in the Downgrade Prediction Model Compared to the SEER Model?

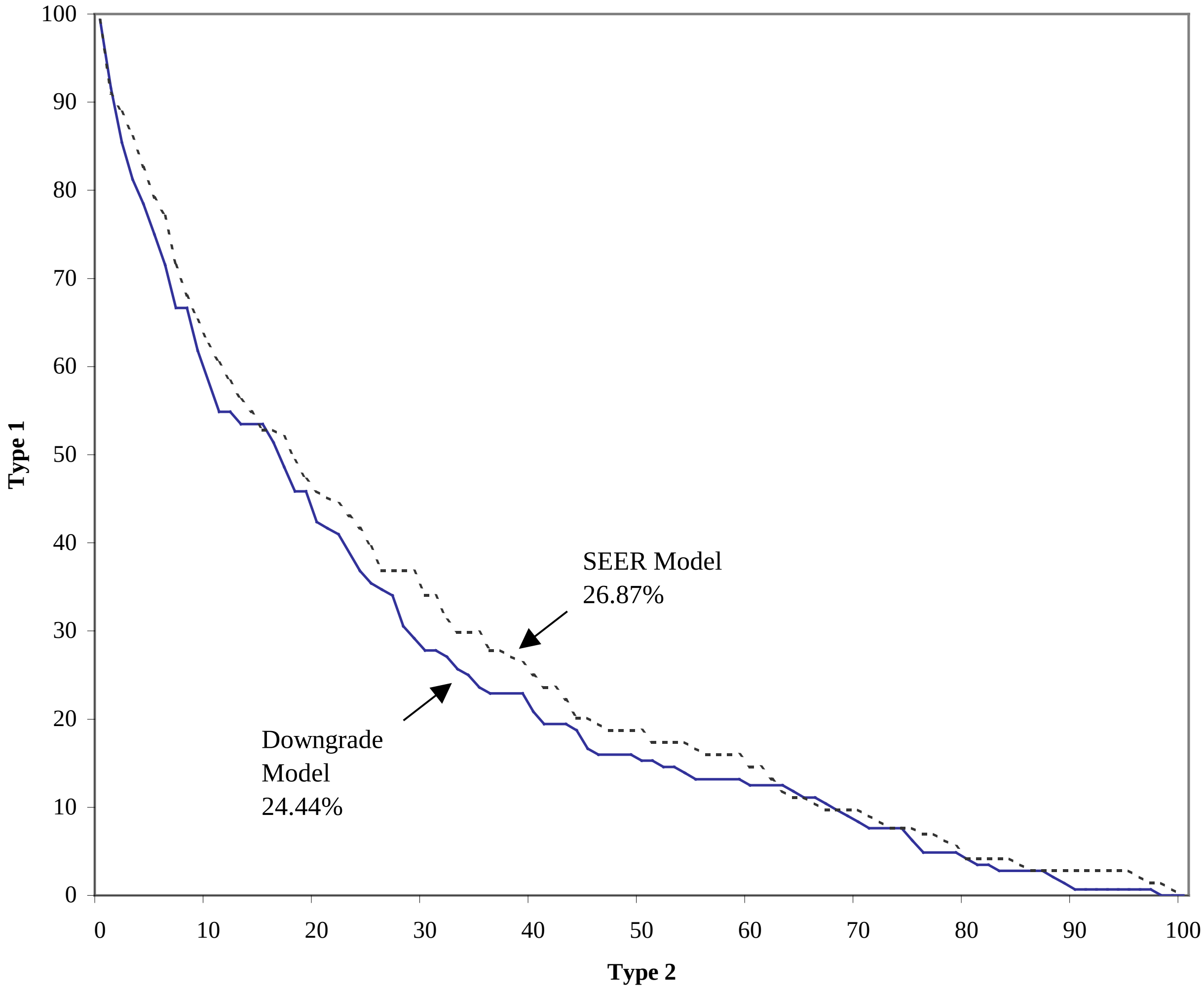
1994 Downgrade Predictions Using Year-end 1992 Data



This figure shows the trade-off between the type-1 error rate and the type-2 error rate. The type-1 error rate is the percentage of banks rated CAMEL-1 or -2 that were subsequently downgraded by supervisors but were not identified by the given model. The type-2 error rate is the percentage of banks rated CAMEL-1 or -2 that were not subsequently downgraded but were misidentified by the given model as a downgrade risk. A desirable early-warning system minimizes the level of type-2 errors for any given level of type-1 errors. This graph shows a similar tradeoff between type-1 and type -2 errors with both models. A convenient way to express the difference between the performance of these models is to calculate the percentage of the area in the figure that is under each curve. Smaller areas are more desirable because they imply simultaneously low levels of both types of errors. In the above graph, the downgrade model slightly edges out the SEER model by 29.49 percent to 30.66 percent.

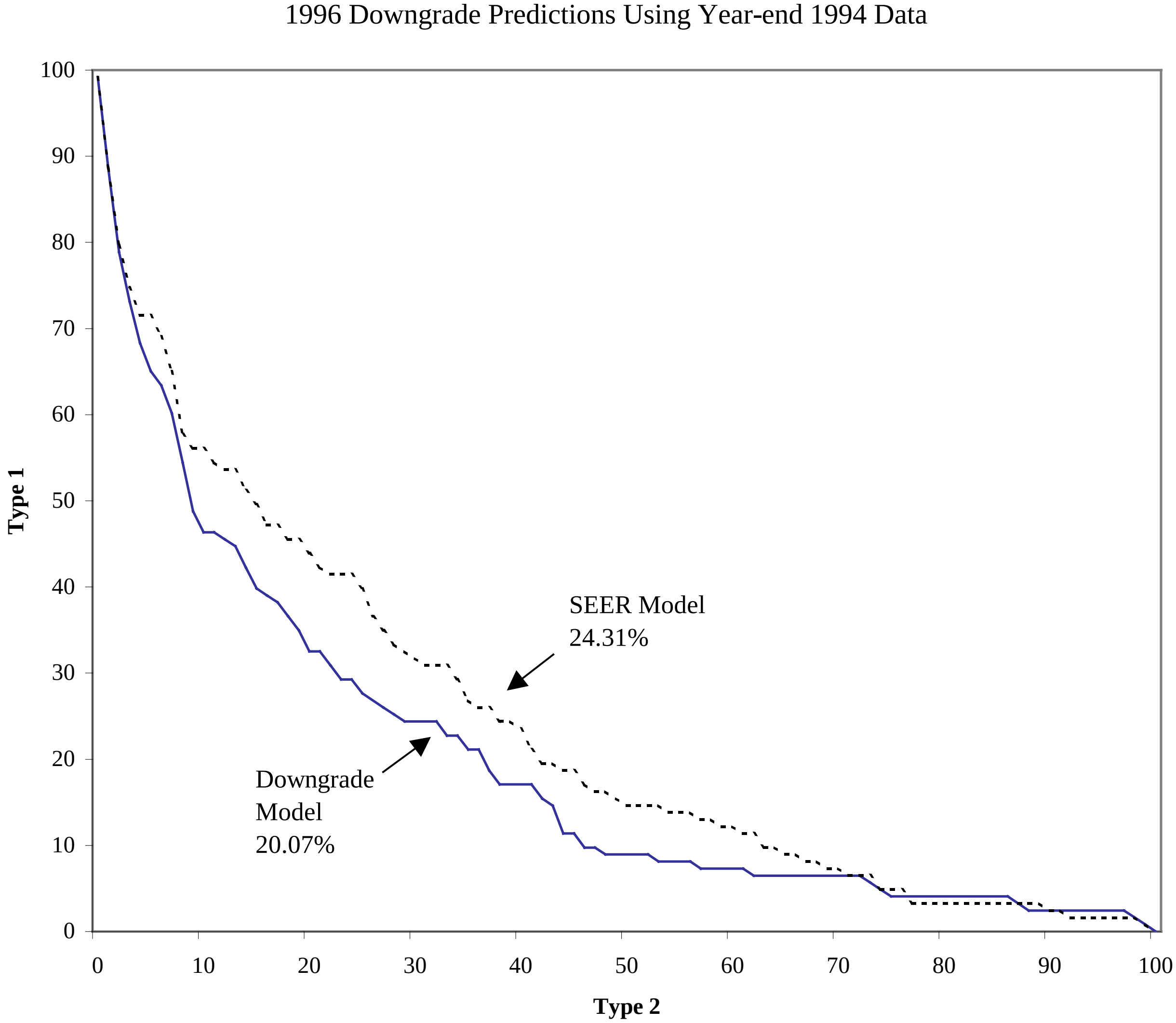
Figure 3: What is the Trade-off Between False Negatives and False Positives in the Downgrade Prediction Model Compared to the SEER Model?

1995 Downgrade Predictions Using Year-end 1993 Data



This figure shows the trade-off between the type-1 error rate and the type-2 error rate. The type-1 error rate is the percentage of banks rated CAMEL-1 or -2 that were subsequently downgraded by supervisors but were not identified by the given model. The type-2 error rate is the percentage of banks rated CAMEL-1 or -2 that were not subsequently downgraded but were misidentified by the given model as a downgrade risk. A desirable early-warning system minimizes the level of type-2 errors for any given level of type-1 errors. This graph shows that for almost all levels of type-1 error, the downgrade model leads to fewer type-2 errors than the SEER failure model. A convenient way to express this difference in performance is to calculate the percentage of the area in the figure that is area under each curve. Smaller areas are more desirable because they imply simultaneously low levels of both types of errors. In the above graph, the downgrade model outperforms the SEER model by 24.44 percent to 26.87 percent.

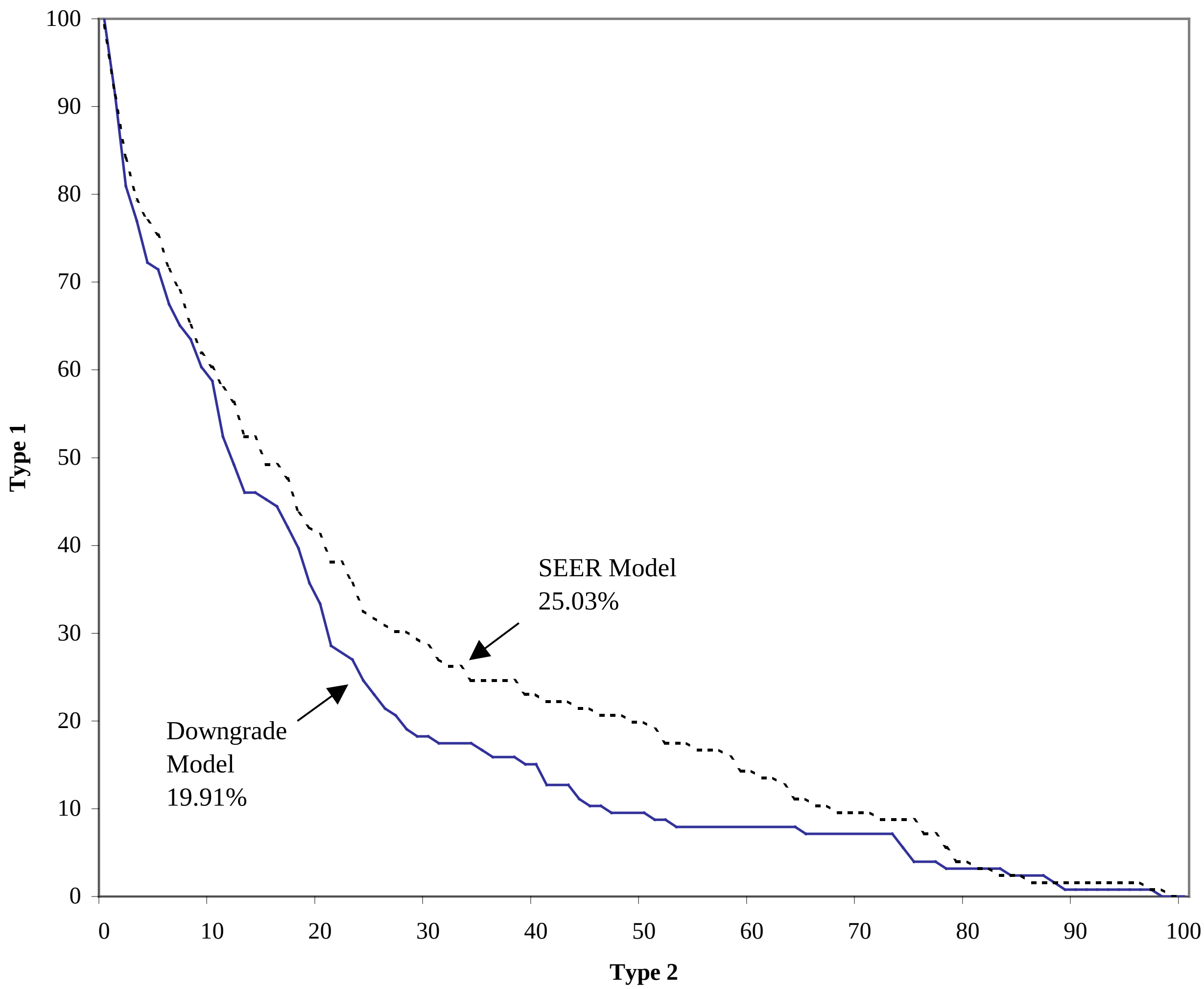
Figure 4: What is the Trade-off Between False Negatives and False Positives in the Downgrade Prediction Model Compared to the SEER Model?



This graph shows that for most levels of type-1error, the downgrade model leads to fewer type-2 errors than the SEER failure model. A convenient way to express this difference in performance is to calculate the percentage of the area in the figure that is under each curve. Smaller areas are more desirable because they imply simultaneously low levels of both types of errors. In the above graph, the downgrade model outperforms the SEER model by 20.07 percent to 24.31 percent.

Figure 5: What is the Trade-off Between False Negatives and False Positives in the Downgrade Prediction Model Compared to the SEER Model?

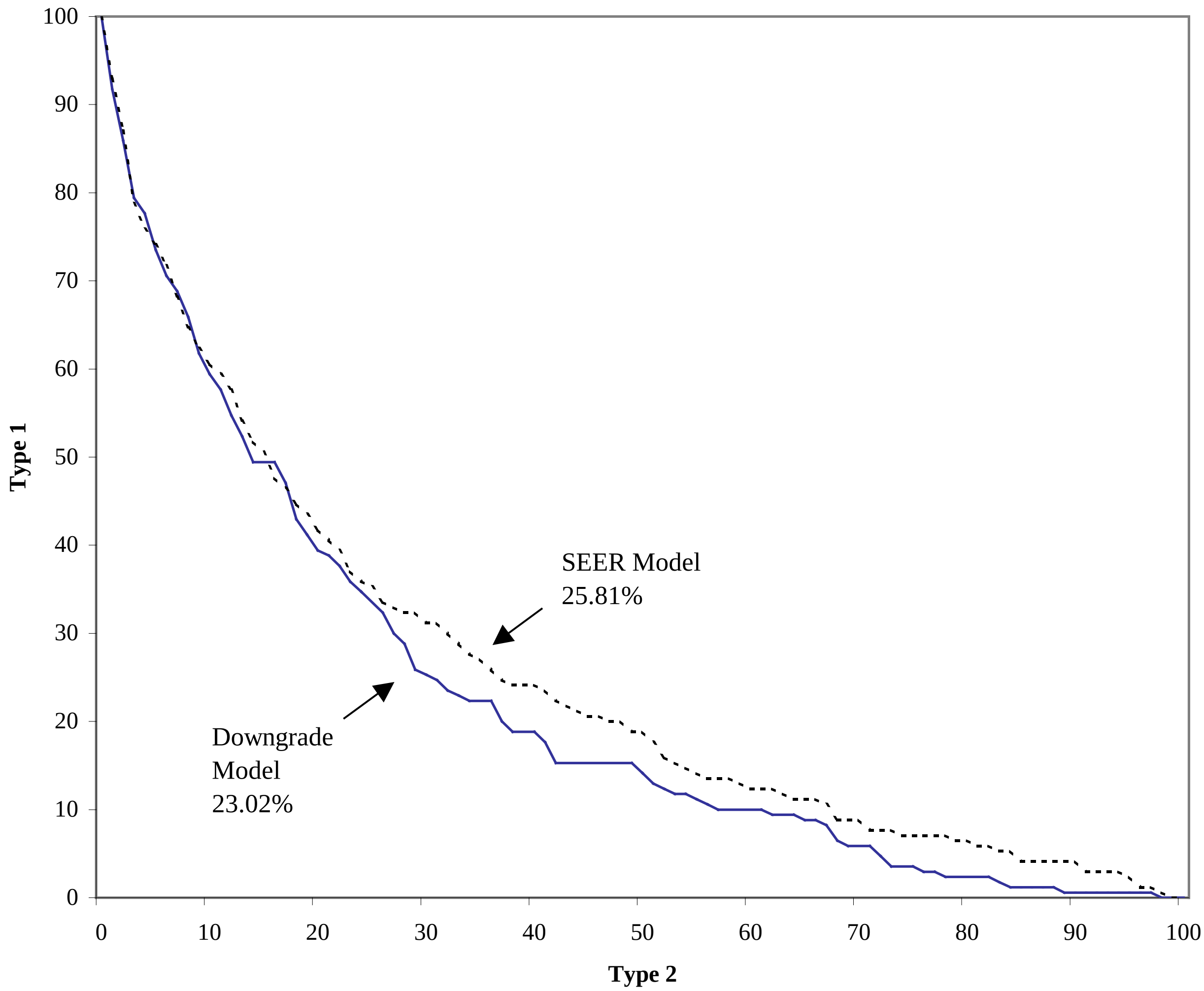
1997 Downgrade Predictions Using Year-end 1995 Data



This graph shows that for almost all levels of type-1 error, the downgrade model leads to fewer type-2 errors than the SEER failure model. A convenient way to express this difference in performance is to calculate the percentage of the area in the figure that is under each curve. Smaller areas are more desirable because they imply simultaneously low levels of both types of errors. In the above graph, the downgrade model outperforms the SEER model by 19.91 percent to 25.03 percent.

Figure 6: What is the Trade-off Between False Negatives and False Positives in the Downgrade Prediction Model Compared to the SEER Model?

1998 Downgrade Predictions Using Year-end 1996 Data



This graph shows that for almost evey level of type I error rate tolerated by supervisors, the downgrade model leads to slightly fewer type-2 errors than the SEER failure model. A convenient way to express this difference in performance is to calculate the percentage of the area in the figure that is under each curve. Smaller areas are more desirable because they imply simultaneously low levels of both types of errors. In the above graph, the downgrade model outperforms the SEER model by 23.02 percent to 25.81 percent.