Using Securities Market Information for Bank Supervisory Monitoring*

John Krainer and Jose A. Lopez
Economic Research Department, Federal Reserve Bank of San Francisco

U.S. bank supervisors conduct comprehensive inspections of bank holding companies and assign them a supervisory rating, known as a BOPEC rating prior to 2005, meant to summarize their overall condition. We develop an empirical model of these BOPEC ratings that combines supervisory and securities market information. Securities market variables, such as stock returns and bond yield spreads, improve the model's in-sample fit. Debt market variables provide more information on supervisory ratings for banks closer to default, while equity market variables provide useful information on ratings for banks further from default. The out-of-sample accuracy of the model with securities market variables is little different from that of a model based on supervisory variables alone. However, the model with securities market information identifies additional ratings downgrades, which are of particular importance to bank supervisors who are concerned with systemic risk and contagion.

JEL Codes: G14, G21.

*The views expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of San Francisco or the Federal Reserve System. We thank Rob Bliss for generously sharing his bank holding company debt database with us. We thank Hyun Song Shin (the editor), an anonymous referee, Fred Furlong, Reint Gropp, and seminar participants at the Bank of England, the Basel Committee's Research Task Force on Banking Supervision, and the Chicago Conference on Bank Competition and Structure for helpful suggestions. We thank Judy Peng Feria, Ryan Stever, and Chishen Wei for research assistance. E-mails: Krainer: john.krainer@sf.frb.org; Lopez: jose.a.lopez@sf.frb.org.

1. Introduction

The most comprehensive tool for banking supervision in the United States is the on-site inspection, where a team of supervisors visits an institution and analyzes its operations in detail. For bank holding companies (BHCs), Federal Reserve examiners assign the institution a rating, known as a BOPEC rating up through 2004, at the conclusion of an inspection. The rating summarizes the examiners' opinion of the BHC's overall financial condition. On-site inspections are conducted on roughly an annual basis. Between BHC on-site inspections, examiners engage in off-site monitoring, which largely consists of analyzing quarterly reports from the institutions in question. However, BHCs that have outstanding public securities are also being monitored by their equity shareholders and debtholders. Since their assessments of BHCs have been found to be timely and accurate in previous studies, the aim of this paper is to investigate the effectiveness of BHC securities market data, both from the equity and debt markets, in off-site monitoring models of BOPEC ratings.[1]

Securities market prices should, in an ideal world, tell supervisors all they need to know about a BHC's condition and its likelihood of failure. In practice, however, there are various frictions that make our questions worthy of empirical research. First, perceptions of possible government support for a struggling BHC, and the depository safety net in general, might reduce investor incentives to monitor. This would potentially decrease the sensitivity of security prices to changes in BHC conditions. Second, since banks specialize in solving problems of asymmetric information, the loans they hold as assets may be difficult for outside investors to value. This problem, like the first, might make security prices less sensitive to changes in asset value. Finally, supervisors have access to information that BHCs are not normally required to disclose to investors, raising questions of whether securities prices can tell supervisors anything they do not already know.

[1]Note that in this paper we focus on supervisory ratings and not defaults, another key supervisory concern. There exists an extensive literature on bank default dating back to Meyer and Pifer (1970), Sinkey (1975), and Pettway and Sinkey (1980).

In this paper, we seek to accomplish three main objectives. First, we investigate the potential contributions of both equity and debt market information to the supervisory monitoring of BHCs using an off-site monitoring model for BOPEC ratings. We measure the contributions of these variables relative to a model based on supervisory data alone.[2]

The second objective of our study is to investigate the cases when information from the equity or debt markets is relatively more useful. We find an asymmetric contribution from these variables for predicting BOPEC rating changes that depends on how close a BHC's assets are to its default point. That is, in a model of BOPEC changes, the magnitude of the coefficients on equity variables is larger for BHCs further from default, while the coefficients on the debt variables are larger for BHCs close to default.

Finally, our third objective is to conduct a set of realistic forecasting exercises aimed at assessing the value of securities market data to bank supervisors over the last banking cycle. We find little evidence of improved forecast performance after incorporating securities market information into the off-site monitoring model. The performance of BOPEC forecasts based on supervisory data alone is not statistically different from that of BOPEC forecasts generated by the model augmented with securities market data. However, we find that while the forecasts are not different in a statistical sense, they are different in an economic sense. The forecasts based on the model incorporating securities market information identify additional BOPEC rating changes, especially downgrades, of publicly traded BHCs that were not identified by a benchmark model based just on supervisory data. Given a supervisory objective function that values early warnings of potential downgrades, this identification of additional correct BOPEC changes could outweigh the cost of additional false signals.

An extensive academic literature regarding the complementarity of supervisory and financial market monitoring of BHCs and their subsidiary banks already exists; see Flannery (1998) for a survey. The majority of the recent studies in the literature have

[2]Throughout the paper, we use the term "supervisory data" to mean data generated by supervisors as part of the BHC quarterly reporting process or as part of the supervisory process.

focused on the various uses of bond market data for monitoring. For example, Flannery and Sorescu (1996) and Covitz, Hancock, and Kwast (2004a, 2004b) study how the sensitivity of bank bond prices changed with the passage of the FDIC Improvement Act, an event that arguably lessened the probability that bondholders would be bailed out in case of a bank failure. DeYoung et al. (2001) find that bank subordinated debt prices do not immediately reflect information generated by supervisors, but that this information does enter prices over several quarters. Evanoff and Wall (2002) find that subordinated debt spreads do as well or better than standard capital ratios at explaining supervisory ratings.

Relatively less attention has been paid to the value of equity market data for supervisory monitoring; for surveys of this work, see Krainer and Lopez (2003, 2004). Even less research has focused on combining equity and debt market data, although there are two notable exceptions. Berger, Davies, and Flannery (2000) examine the timeliness and accuracy of supervisory and market assessments of the condition of large BHCs. They find that, after accounting for market assessments, balance sheet variables do not contribute substantially to the modeling of future indicators of BHC performance, such as changes in nonperforming loans, book equity capital, and return on assets. Their findings suggest that supervisors, bond market participants, and equity market participants produce complementary but different information on BHC performance.

Gropp, Vesala, and Vulpes (2004, 2006) examine the ability of equity market variables and subordinated bond spreads for European banks to signal changes in their agency ratings.[3] Using ordered logit models at several horizons and a proportional hazard model, they find that both equity-based measures of distance to default and subordinated debt spreads are useful for detecting changes in bank ratings. Interestingly, they find that the distance-to-default measure performs less well closer to default and that subordinated debt spreads seem to have signal value only close to default. The authors argue that their empirical results provide support for the use of securities market information in supervisors' early-warning models.

[3]See also Gropp and Richards (2001).

Our paper differs from the prior studies in several important ways. As in Berger, Davies, and Flannery (2000), we use data on U.S. bank holding company condition, but our analysis focuses squarely on supervisory ratings as the key measure of BHC condition. Thus, our analysis addresses the benefits of using market information to predict variables that supervisors care most about. Additionally, we extend their in-sample analysis to an out-of-sample forecasting framework that mimics the way supervisors might actually use securities market data as a supervisory tool. Finally, though we find a limited use of securities market information for forecasting BHC condition, we establish the conditions at the BHC under which equity and debt market signals can be useful. Our results provide clear empirical reasoning for the Berger, Davies, and Flannery (2000) finding that supervisory and debt-related assessments of BHCs are related. Thus, our results are a refinement of their finding that supervisory and equity market indicators are not interrelated; i.e., we find that abnormal BHC stock returns do provide useful information with respect to changes in supervisory assessments when BHCs are far from their default points.

The paper proceeds as follows. In section 2, we provide a brief overview of the supervisory process for BHCs in the United States. In section 3, we estimate our proposed BOPEC off-site monitoring model (BOM) using both supervisory and securities market variables. We also examine the differential impact of the securities market variables based on the BHCs' relative distances from their default points. In section 4, we examine the model's out-of-sample performance using a statistical and a supervisory objective function. Section 5 concludes.

2. The U.S. Supervisory Process

The Federal Reserve is the supervisor of bank holding companies in the United States. Full-scope, on-site inspections of BHCs are a key element of this supervisory process. These inspections are generally conducted on an annual basis, particularly for the case of large and complex BHCs.[4] Limited and targeted inspections that may or

[4]A complex BHC is defined as one with material credit-extending nonbank subsidiaries or debt outstanding to the general public. See DeFerrari and Palmer

may not be conducted on-site are also carried out. In this paper, we focus on full-scope, on-site inspections since they provide the most comprehensive supervisory assessments of BHCs.

At the conclusion of an inspection, supervisors assign the BHC a numerical rating, which prior to 2005 was called a composite BOPEC rating, that summarizes their opinion of the BHC's overall health and financial condition.[5] The BOPEC acronym stands for the five key areas of supervisory concern: the condition of the BHC's Bank subsidiaries, Other nonbank subsidiaries, Parent company, Earnings, and Capital adequacy. A BOPEC rating of 1 is the best rating, with a rating of 5 being the worst rating short of closure. A rating of 1 or 2 indicates that the BHC is not considered to be of supervisory concern. Note that BOPEC ratings, as well as all other inspection materials, are confidential and are not made publicly available.

Between on-site inspections, when private supervisory information cannot be gathered as readily, supervisors monitor BHCs using quarterly regulatory reports filed by BHCs and their subsidiary banks. In addition, the supervisory CAMELS ratings assigned to banks within the holding company are used for monitoring the parent BHCs. As with BOPEC ratings, CAMELS ratings are confidential ratings that are assigned after a bank examination. The acronym refers to the six key areas of concern: the bank's Capital adequacy, Asset quality, Management, Earnings, Liquidity, and Sensitivity to risk. The composite CAMELS rating also ranges in integer value from 1 to 5 in decreasing order (i.e., banks that perform best are assigned a rating of 1). Since the condition of a BHC is closely related to the condition of its subsidiary banks, the off-site BHC surveillance program includes monitoring recently assigned CAMELS ratings.

As with on-site BHC inspections, on-site bank examinations occur at approximately a yearly frequency, which is long enough for the gathered supervisory information to decay and become less

(2001) for an overview of the supervisory process for large, complex banking organizations.

[5]Note that in January 2005, the BOPEC rating system was replaced by the RFI/C(D) rating system; see Board of Governors of the Federal Reserve System (2004). For an international survey of supervisory bank rating systems, see Sahajwala and van den Bergh (2000).

representative of the bank's condition.[6] To address this issue, the Federal Reserve instituted an off-site monitoring system for banks, known as the System for Estimating Examiner Ratings (SEER), in 1993. The SEER system actually consists of two separate models that forecast bank failures over a two-year horizon as well as bank CAMELS ratings for the next quarter. The model that we are most interested in here is the latter, which is an ordered logit model with five categories corresponding to the five possible values of the CAMELS rating. The model is estimated every quarter in order to reflect the most recent relationship between the selected financial ratios and the two most recent quarters of CAMELS ratings. Significant changes in a bank's CAMELS rating as forecasted by the SEER model could be sufficient to warrant closer monitoring of the bank. The off-site BHC surveillance program also explicitly monitors the SEER model's forecasted CAMELS ratings.

2.1 The BOPEC Ratings Sample

The core database for our analysis is the set of supervisory BOPEC ratings assigned after on-site, full-scope inspections between the first quarter of 1990 and the second quarter of 1998. The sample endpoint is dictated by the availability of the bond data set.[7] Our sample of BOPEC ratings is further refined to include only those inspections of top-tier BHCs with identifiable lead banks, four quarters of prior supervisory data, and at least one prior BOPEC rating. We focus on top-tier BHCs since they are typically the legal entities within the banking group that issues publicly traded equity. The lead-bank designation is often provided by banks in their regulatory filings. When such self-reporting is not available, we assign the lead-bank designation to the largest bank within the group. We need the BHCs in our sample to have identifiable lead banks in order to directly link their BOPEC ratings to their lead bank's CAMELS

[6]See Cole and Gunther (1998) as well as Hirtle and Lopez (1999) for further discussion of this issue.

[7]We are grateful to Rob Bliss for sharing his BHC bond database with us. A complete description of the database is presented in Bliss and Flannery (2001). The last quarter of bond data is the first quarter of 1998, which aligns with the second quarter of 1998 in our modeling framework.

ratings.[8] Finally, we require each BHC to have at least four quarters of supervisory data and a lagged BOPEC rating in order to avoid issues regarding de novo BHCs and new BHCs arising from mergers. In addition, four quarters of supervisory data are required to calculate certain explanatory variables for the model described below.

Table 1 summarizes our sample of 3,010 BOPEC ratings assigned to 1,034 unique entities. Almost 65 percent of the BHCs in the sample are relatively small, with less than $1 billion in total assets. Slightly more inspections occurred in the first half of the sample than in the second half, reflecting consolidation in the U.S. banking sector. For publicly traded BHCs, there are 1,291 BOPEC assignments corresponding to 363 unique entities. Note that public BHCs are generally larger than privately held BHCs, with a greater percentage having total assets ranging between $1 billion and $100 billion. Of the 41 BOPECs assigned to the largest BHCs, 39 are of public BHCs. With respect to BHCs with public debt outstanding, this subsample contains 309 BOPEC ratings corresponding to 63 unique BHCs. Again, these BHCs are typically larger than those in the full sample, with almost all BHCs having between $1 billion and $100 billion in assets. Finally, there are 283 BOPEC ratings corresponding to 58 unique BHCs that have both publicly traded equity and debt outstanding during the sample period.

Tables 2A and 2B present the distribution of BOPEC ratings assigned in each year for all BHCs, as well as for BHCs with both publicly traded equity and publicly traded bonds. The distributions in the two tables share many common features. The majority of the ratings fall in the upper two categories, indicating that a BHC's financial condition and risk profile are of little supervisory concern. For the full sample, while the distribution of ratings fluctuates over time, the percentage of ratings in the top two categories never falls below 63 percent. The maximum value is 96 percent in 1998. In contrast, for larger BHCs with both public equity and debt, the percentage of ratings in the upper two categories ranges from 47

[8]Note that this restriction does not imply that we limited the sample to single-bank BHCs. We simply focus on the CAMELS rating for a BHC's lead bank, whether self-identified or identified by asset size.

Table 1. Asset Size Characteristics of the BHCs in the BOPEC Sample

	1990–98	**1990–94**	**1995–98**
Total # Ratings	3,010	1,735	1,275
Asset Size:			
Assets > \$100b	41	13	28
\$1b < Assets < \$100b	1,019	594	425
Assets < \$1b	1,950	1,128	822
Total # Ratings, Public BHCs	1,291	741	550
Asset Size:			
Assets > \$100b	39	13	26
\$1b < Assets < \$100b	807	487	320
Assets < \$1b	445	241	204
Total # Ratings, BHCs with Public Debt	309	174	135
Asset Size:			
Assets > \$100b	37	11	26
\$1b < Assets < \$100b	270	163	107
Assets < \$1b	2	0	2
Total # Ratings, Public BHCs with Public Debt	283	163	120
Asset Size:			
Assets > \$100b	36	11	25
\$1b < Assets < \$100b	247	152	95
Assets < \$1b	0	0	0

Note: The data sample spans the period from the first quarter of 1990 to the second quarter of 1998. The definition of a bank holding company (BHC) used in this table is the definition used in constructing our data set; i.e., a top-tier BHC with an identifiable lead bank, four quarters of available regulatory reporting data, and a lagged BOPEC rating. Public debt here refers to publicly traded bonds as listed in the Warga/Lehmann data set and used by Bliss and Flannery (2001).

percent in 1991 to 100 percent from 1996 onward. Note that there are relatively few inspections culminating in a BOPEC rating of 4 or worse, with the exception of the early 1990s. This likely reflects the fact that both supervisors and bankers actively try to prevent this

Table 2A. All BOPEC Ratings in the Sample

	BOPEC Rating % of Total			
	1	**2**	**3**	**4–5**
1990	16%	52%	21%	10%
1991	16%	47%	25%	12%
1992	15%	52%	20%	14%
1993	24%	55%	14%	7%
1994	34%	53%	8%	5%
1995	36%	52%	8%	5%
1996	47%	47%	5%	1%
1997	47%	48%	4%	0%
1998	47%	49%	3%	0%
Total	31%	51%	12%	6%

Note: The data sample spans the period from the first quarter of 1990 to the second quarter of 1998.

Table 2B. All BOPEC Ratings for BHCs with Public Equity and Bonds in the Sample

	BOPEC Rating % of Total			
	1	**2**	**3**	**4–5**
1990	21%	55%	17%	7%
1991	12%	35%	31%	23%
1992	18%	41%	26%	15%
1993	24%	68%	3%	6%
1994	45%	50%	5%	0%
1995	44%	56%	0%	0%
1996	60%	40%	0%	0%
1997	58%	42%	0%	0%
1998	68%	32%	0%	0%
Total	38%	48%	9%	5%

Note: The data sample spans the period from the first quarter of 1990 to the second quarter of 1998.

outcome, as well as the general good health of the banking sector in the mid and late 1990s.

3. Multivariate Analysis Using the Ordered Logit Model

Our proposed BOPEC off-site monitoring (BOM) model is an ordered logit model and is similar in structure to the SEER model for CAMELS ratings; see Krainer and Lopez (2003, 2004) for further details.[9] The model assumes that the supervisory rating assigned to BHC i in quarter t is a continuous variable, denoted BP^*_{it}, for which only discrete values, denoted BP_{it}, are observed. Recall that lower values of BP^*_{it} and BP_{it} correspond to better supervisory ratings. We model the continuous variable as

$$\mathrm{BP}^*_{it} = \beta \mathrm{x}_{it-2} + \varepsilon_{it}. \tag{1}$$

In this specification, x_{it-2} is a $(\mathrm{k} \times 1)$ vector of supervisory variables unique to BHC i observed two quarters prior to the BOPEC assignment. We chose to lag the supervisory variables by two quarters because, in real time, these are often the most recent data available at the time of inspection.[10]

For firms with publicly traded securities, we augment equation (1) with securities market data, which is denoted as z_{it-1}; i.e.,

$$\mathrm{BP}^*_{it} = \beta \mathrm{x}_{it-2} + \pi \mathrm{z}_{it-1} + \varepsilon_{it}. \tag{2}$$

The vector z_{it-1} contains securities market variables—equity market variables, debt market variables, or both depending on the model—corresponding to BHC i at time $t-1$, one quarter prior to the BOPEC assignment. The supervisory variables and the securities market variables enter into the model with different lags because securities market data are available on a more timely basis than are the supervisory variables that we use. The error term ε_{it} has a standard logistic distribution. Note that we estimate four BOM

[9]Recall that Gropp, Vesala, and Vulpes (2004, 2006) used logit and proportional hazard models, which are better suited for examination of the timing and determinants of banks' Fitch/IBCA ratings changes.

[10]See Gunther and Moore (2003) for further discussion.

models corresponding to the four alternative z_{it-1} vectors using BOPEC samples based on the availability of securities market variables. For example, we estimate the core model without securities market variables (i.e., $z_{it-1} = 0$) on all BOPEC ratings, but we estimate the BOM model using both sets of market variables only on the BOPEC sample for which both sets of variables are available.

3.1 Supervisory Variables

The choice of which supervisory variables to include in x_{it-2} is challenging. No simple behavioral models exist of how supervisors assign BOPEC ratings. As mentioned, there are more than 800 variables at the supervisors' disposal for this purpose. For this study, we selected nine explanatory variables that were judged to be reasonable proxies for the five components of the BOPEC rating. In addition, we included lagged BOPEC ratings in the specification to control for ratings persistence. As in Krainer and Lopez (2003, 2004), we chose a parsimonious specification in the hopes of generating reasonable out-of-sample forecasts.

Our supervisory variables are summarized in table 3. The first variable is the natural log of total BHC assets, which is our control variable for BHC size. The next four variables are used to capture the supervisory concerns regarding the BHC's bank subsidiaries, as summarized in the "B" component of the rating. The second variable is the CAMELS rating of the BHC's lead bank. The third variable is the ratio of the BHC's nonperforming loans, nonaccrual loans, and other real estate owned to its total assets. This variable proxies for the health and performance of the BHC's loan portfolio. The fourth variable is the ratio of the BHC's allowances for losses on loans and leases to its total loans, another proxy for the health and performance of the BHC's loan portfolio.

The fifth variable is an indicator of whether the BHC has a section 20 subsidiary, which is a subsidiary that can engage in securities activities that commercial banks were generally not permitted to engage in before the passage of the Gramm-Leach-Bliley Act of 1999. This variable is a proxy for the scale of the BHC's nonbank activities and thus speaks to the "O" component of the BOPEC rating. We also include as the sixth variable the ratio of a BHC's trading assets

Table 3. Summary Statistics for Financial Statement and Supervisory Variables

	Mean	Std. Dev.	25th Pctile.	Median	75th Pctile.
Assets ($m)	$6,336	$23,700	$250	$493	$2,068
CAMELS Rating	1.94	0.80	1	2	3
Nonperforming Loans/Assets	1.97%	1.87%	0.87%	1.47%	2.41%
Allowances for Loan Losses/ Assets	0.41%	0.69%	0.09%	0.21%	0.44%
Section 20 Subsidiary	0.04	0.19	0.00	0.00	0.00
Trading Assets/ Assets	1.10%	42.27%	0.0%	0.0%	0.0%
Double Leverage	55.24%	108.34%	7.29%	43.27%	98.21%
Return on Average Assets	0.82%	0.97%	0.66%	0.98%	1.22%
Equity Capital/Assets	8.18%	2.47%	6.71%	7.88%	9.26%

to its total assets as a proxy of its nonbanking activities, whether conducted in banking or nonbanking subsidiaries.[11]

The seventh variable is the so-called "double leverage" ratio between the BHC and its lead bank, which is the ratio of the lead bank's equity capital to that of the parent's equity capital. This variable provides a measure of the soundness of the parent BHC, indicating the extent to which the parent's equity capital can be used to buffer against damage to the lead bank's equity capital. We use this variable as a proxy for the condition of the parent BHC as summarized in the "P" component of the BOPEC rating. The eighth variable is the BHC's return on average assets (ROAA), defined as the ratio of the four-quarter average of the BHC's net income to the four-quarter average of its assets. This variable is used to proxy for

[11]Note that the trading assets variable as currently reported first became available in the first quarter of 1995. Before then, we proxy for BHC trading assets using the sum of the self-reported replacement cost of interest rate and foreign exchange derivative contracts.

the "E" component of the BOPEC rating. The ninth variable is the BHC's ratio of equity capital to its total assets. This variable is used to proxy for the "C" component of the BOPEC rating.[12]

3.1.1 Equity Market Variables

The two equity market variables used in this study are based on BHC stock returns over the quarter prior to a BOPEC assignment. The motivation behind using equity market data lies in the conjecture that there is some agreement between BHC stock market investors and supervisors on what determines healthy financial condition. Stock market investors clearly are not trying to forecast BOPEC ratings. However, if the same financial developments that lead to supervisory rating changes also lead to changes in expected stock returns, it is possible that supervisors could use stock market signals as an additional off-site monitoring signal.

As per Campbell, Lo, and Mackinlay (1997), we decompose monthly BHC stock returns, denoted as R_{it}, before a BOPEC assignment into a systematic return, denoted SR_{it}, and an idiosyncratic, or abnormal, return, denoted AR_{it}. The decomposition is based on a two-factor market model; i.e.,

$$\mathrm{R}_{it} = \alpha + \beta_1 \mathrm{R}_{mt} + \beta_2 \mathrm{f}_t + \nu_{it}, \tag{3}$$

where R_{mt} is the monthly return on the CRSP value-weighted index, f_t is the monthly change in the federal funds rate, and ν_{it} is a normally distributed error term. For each BOPEC assignment in our sample, the model's parameters are estimated using monthly data over a period of at least two years leading up to one year prior to the assignment of the BOPEC rating. When available, we used up to five years of monthly data for the estimation window. In the twelve-month event window leading up to the assignment, the systematic portion of the return, SR_{it}, is calculated as $\mathrm{SR}_{it} = \hat{\alpha} + \hat{\beta}_1 \mathrm{R}_{mt} + \hat{\beta}_2 \mathrm{f}_t$. The corresponding abnormal return, AR_{it}, is simply the difference between the realized return and the systematic return, $\mathrm{AR}_{it} = \mathrm{R}_{it} - \mathrm{SR}_{it}$. The stock market variables

[12] A variety of capital measures have been used in previous studies, such as Estrella, Park, and Peristiani (2000) and Evanoff and Wall (2001). We chose a simple measure to facilitate comparison over the entire period.

are constructed to detect changes in BHC condition that occur between inspections and are relevant to an eventual BOPEC assignment but are not yet embedded in the available supervisory data. Thus, the stock market variables are used in event-study fashion, where we cumulate systematic and abnormal returns over the three-month window between quarter $t-2$, when the latest supervisory data are available, and quarter $t-1$, the quarter before the assignment.

In the empirical work to follow, the quarterly cumulative return at time $t-1$, CR_{it-1}, is the actual return between month $t-6$ and month $t-3$. As an example, consider a BOPEC assignment in the last quarter of the year. The cumulative return is the sum of the three monthly returns from July to September. The systematic and idiosyncratic returns are formed in the same way. To permit comparison across BHCs and across time, we standardize the cumulative returns using the estimated standard errors for CAR_{it-1}. The standardized form of these variables is

$$\frac{\mathrm{CR}_{it-1}}{\sqrt{\mathrm{var}(\mathrm{CAR}_{it-1})}} = \frac{\mathrm{CSR}_{it-1}}{\sqrt{\mathrm{var}(\mathrm{CAR}_{it-1})}} + \frac{\mathrm{CAR}_{it-1}}{\sqrt{\mathrm{var}(\mathrm{CAR}_{it-1})}}, \tag{4}$$

or equivalently, $\mathrm{SCR}_{it-1} = \mathrm{SCSR}_{it-1} + \mathrm{SCAR}_{it-1}$.

BHC stock price changes that are large in magnitude with respect to general market activity may signal changes in condition that will eventually lead to a ratings change. The SCAR variable is designed specifically for identifying which stock price changes are large based on our normality assumption. However, relying exclusively on SCARs for market signals may cause us to miss important information available from the broader stock market. For example, an economy-wide shock that lowers returns for all stocks might not translate into abnormally negative returns for any particular BHC but could very well be an early indicator of changes in all supervisory ratings. To address this concern, we include the SCSR variable in our regressions. Hence, $\mathrm{z}_{Eit} = [\mathrm{SCSR}_{it}, \mathrm{SCAR}_{it}]$.

It is important to note that there are circumstances where stock return data, especially if not normalized in some way, could be misleading from a supervisory perspective. For example, a BHC close to default has incentive to increase its risk profile, and the call option feature of the equity claim should cause the stock price to rise. In this scenario, supervisors could potentially mistake the stock market signal of greater risk for a signal of improved condition. To address this concern, Gropp, Vesala, and Vulpes (2004) advocate using an equity-based distance-to-default measure that accounts for this increased risk. In contrast, we focus on the cumulative return measures described above, because these variables are more appropriate for forecasting exercises, where market signals are assessed not just by their size but by their relative precision (i.e., their standard errors).

3.1.2 Debt Market Variables

The debt market variable used in this study is adjusted bond yields from the Warga/Lehmann Brothers Corporate Bond Database, as used by Bliss and Flannery (2001) (henceforth, BF). Note that this database includes both subordinated and nonsubordinated BHC debt.[13] There are two issues that need to be confronted before using the bond data. First, in cases where a BHC has multiple outstanding bonds, it is necessary to compress their market signals into a single observation. Following BF, we use weighted-average BHC bond yields, where the weights correspond to the size of the issue relative to the BHC's total amount of bonds outstanding in the quarter. Second, as with the equity market variables, we would like to have some measure of what constitutes an abnormal change in bond yield. Here, we follow the BF procedure of computing debt spreads from bond price indices based on eleven Moody's ratings categories and three term-to-maturity categories.[14] The BF indices allow us to study debt yields relative to an index of similar bonds drawn from all industries. We also adjusted the BHC yield spreads to account for their last assigned BOPEC ratings.

To summarize, for BHC i with BOPEC rating j at time t, we define the yield on a bond (or a weighted average of several bonds) with terms k (i.e., maturity and Moody's rating) as y_{ijkt}. We then

[13]For a discussion of the market for BHC subordinated debt in the United States, see Board of Governors of the Federal Reserve System (1999), Hancock and Kwast (2001), Basel Committee on Banking Supervision (2003), and Goyal (2005).

[14]The "+" or "−" qualifiers attached to the basic rating definitions are suppressed. The maturity buckets are less than five years, five to ten years, and greater than ten years.

constructed the yield spread s_{ijkt} relative to the corresponding BF based on terms k; i.e.,

$$\mathrm{s}_{ijkt} = \mathrm{y}_{ijkt} - \overline{\mathrm{y}}_{kt}, \tag{5}$$

where $\overline{\mathrm{y}}_{\mathrm{kt}}$ is the yield on an index of like-termed bonds. We then further adjusted the yield spread to account for the BHC's last assigned BOPEC rating j; i.e.,

$$\mathrm{d}_{ijkt} = \mathrm{s}_{ijkt} - \overline{\mathrm{s}}_{jt}, \tag{6}$$

where $\overline{\mathrm{s}}_{jt}$ is the median yield spread for all BHCs with BOPEC rating j at time t and publicly traded debt. In our empirical work, we found that these adjusted yield spreads appear to have more predictive power than the yield spreads based just on the BF indices; hence, $\mathrm{z}_{Dit-1} = \mathrm{d}_{ijkt-1}$.

3.2 Empirical Results

We estimate the BOM model over the four samples of BOPEC ratings. These four samples are the sample consisting of BOPEC ratings, the sample of ratings assigned to BHCs with publicly traded equity, the sample of ratings assigned to BHCs with publicly traded debt, and the sample of ratings assigned with both types of securities. The sample sizes are 3,010; 1,291; 309; and 283 observations, respectively. Note that the sample size decreases by more than a factor of ten from the largest to the smallest samples. The results are presented in tables 4A and 4B.

The basic results for the core model containing just supervisory variables are similar to those reported in Krainer and Lopez (2004). The lagged BOPEC rating and the current bank CAMELS rating are important drivers of the rating process; higher (or worse) values of these ratings lead to higher BOPEC ratings. Higher total assets are correlated with lower (i.e., better) ratings. High levels of problem loans and allowances tend to lead to worse BOPEC ratings. High levels of ROA and capital tend to lead to better BOPEC ratings.[15]

[15] We conducted two important robustness checks for this sample. First, regarding parameter stability across the sample period, we split the sample period in

Table 4A. BOM Model Estimation Results

	Full Sample		Equity Sample	
	Coefficients	**p-value**	**Coefficients**	**p-value**
Lagged BOPEC	1.292*	0.00	1.551*	0.00
CAMELS	1.223*	0.00	0.685*	0.00
Total Assets	−0.247*	0.00	−0.182*	0.00
Problem Loans	48.050*	0.00	50.208*	0.00
Allowances	56.676*	0.00	59.315*	0.01
Trading Assets	0.004	0.49	−1.502	0.42
Section 20 Sub.	1.819*	0.00	0.272	0.36
Double Leverage	0.054	0.25	−0.282	0.16
ROA	−1.015*	0.00	−0.567	0.06
Equity Capital	−22.103*	0.00	−45.280*	0.00
SCSR	–	–	−0.831*	0.00
SCAR	–	–	−0.523*	0.00
Adj. Yield Spread	–	–	–	–
Observations	3,010		1,291	
Pseudo R^2	0.47		0.48	

Note: The model estimated here across the four subsamples is summarized in equation (1). The results for the interacted supervisory variables and fixed effects are not reported, in order to conserve space. The models contained fixed effects for BHCs with public equity, BHCs with public equity for which SCAR and SCSR variables could not be properly calculated, and BHCs with public debt. The sample period range is 1990:Q1 to 1998:Q2. The model was estimated using robust standard errors and adjusting for clustered observations based on unique BHCs. A * denotes significance at the 5 percent level.

half into a "banking crisis" subsample (1990:Q1 to 1994:Q4) and a "banking recovery" subsample (1994:Q2 to 1998:Q2). The estimated coefficients in the core and extended models did not change much over the two subsamples. All coefficient estimates that were significantly different from zero in the full sample remained so and with the same sign in the subsamples. Second, to account for possible BHC merger effects during the period, we dropped all observations where the BHC had been involved in a merger, either as an acquirer or as an acquired institution, one year prior to its BOPEC assignment. This resulted in a loss of 771 observations, or about 26 percent of our sample. The coefficient estimates are very similar to those reported in tables 4A and 4B. Detailed results from these robustness checks are available from the authors upon request.

Table 4B. BOM Model Estimation Results

	Debt Sample		Equity and Debt Sample	
	Coefficients	***p*-value**	**Coefficients**	***p*-value**
Lagged BOPEC	2.144*	0.00	2.587*	0.00
CAMELS	0.427	0.08	0.465	0.11
Total Assets	0.119	0.54	0.092	0.68
Problem Loans	80.497*	0.00	95.568*	0.00
Allowances	2.313	0.95	14.104	0.63
Trading Assets	−5.865*	0.01	−4.803*	0.05
Section 20 Sub.	0.225	0.57	0.190	0.66
Double Leverage	−0.529	0.26	−0.224	0.63
ROA	−0.840	0.14	−0.509	0.31
Equity Capital	−60.633	0.00	−69.942*	0.00
SCSR	–	–	−0.552*	0.04
SCAR	–	–	−0.972*	0.00
Adj. Yield Spread	2.688*	0.00	2.382	0.07
Observations	309		283	
Pseudo R^2	0.48		0.58	

Note: The model estimated here across the four subsamples is summarized in equation (1). The results for the interacted supervisory variables and fixed effects are not reported, in order to conserve space. The models contained fixed effects for BHCs with public equity, BHCs with public equity for which SCAR and SCSR variables could not be properly calculated, and BHCs with public debt. The sample period range is 1990:Q1 to 1998:Q2. The model was estimated using robust standard errors and adjusting for clustered observations based on unique BHCs. A * denotes significance at the 5 percent level.

For the sample associated with publicly traded BHCs, the empirical results for the supervisory variables are practically identical. The two equity market variables have negative signs as expected; i.e., increases in either systemic or idiosyncratic stock returns tend to be associated with better BHC performance and better (i.e., lower) BOPEC ratings.

For the sample associated with BHCs with publicly traded debt, the empirical results for the supervisory variables change somewhat. Total assets and allowances are no longer significant drivers of BOPEC ratings, while trading assets are significant, with higher values leading to better ratings. The *p*-values on current bank CAMELS ratings and ROA increase beyond the 5 percent significance level. However, lagged BOPEC ratings, problem loans, and equity capital remain important explanatory variables. For the adjusted yield spread variable, the positive sign is also in line with expectations; i.e., higher yield spreads relative to the corresponding yield index are associated with worse BHC performance and worse (i.e., higher) BOPEC ratings.[16] The empirical results of the sample for BHCs with both types of publicly traded securities are quite similar to those of the debt sample. Notably, the adjusted yield spread has a positive coefficient with a p-value of 7 percent, just outside of our defined significance level.

These in-sample results clearly suggest that securities markets provide information complementary to that generated by supervisors. Interestingly, the results for BHCs with both public equity and debt seem to suggest that equity market variables may be more useful than debt market variables. In the next subsection, we examine more directly the relative importance of these two sources of market information.

3.3 The Relative Importance of Equity and Debt Market Information

As noted in the introduction, a primary motivation for using both equity and debt market data in a supervisory monitoring model is that no single information source is likely to dominate the other in all states of the world. How, then, might these two types of securities market information differ? In which states of the world are the different sources of information most useful? We might expect that the residual claim feature of equity implies that equity market investors would be good at predicting BOPEC rating changes when BHC asset values are relatively far from the default point. For

[16]We estimated versions of the models above that allowed the coefficients on securities market variables to differ according to whether the market might perceive the institution to be "too big to fail." For this analysis, we classified a BHC as possibly being too big to fail if it was one of the five largest BHCs by assets in a given year. We found that equity market prices have the same explanatory value for BOPEC ratings for this class of BHCs as for the rest of the sample. Debt prices for the largest BHCs were considerably less useful for signaling changes in condition, although this effect was not measured precisely. Results from this robustness check are available from the authors upon request.

such total asset values, changes in value correspond one-to-one with changes in equity value. Debt market investors, by contrast, might be more likely to predict BOPEC rating changes when asset values are relatively close to or below the default point, as that range for asset values is where bondholders are most at risk for taking losses.

To examine the relationship between securities market signals and how close a BHC might be to default, we explicitly calculate a default point for each public BHC at each point in time and examine how its interaction with the securities market variables affects the supervisory ratings. For this exercise, we use the Ronn-Verma (1986) model for estimating the value of a firm's assets. Given assumptions on the stochastic process describing changes in asset value, the firm's equity is modeled as a call option on those assets. This framework gives rise to a distance-to-default (DTD) measure, which is the difference between the estimated market value of the assets and the equity, scaled by estimated standard deviation of the rate of return on the assets. Note that the construction of this measure restricts our sample to public BHCs.

Working again with ordered logit models within the BOM framework, we create indicator variables for whether a BHC's DTD measure is within a certain percentile of the overall sample distribution of DTD measures. We constructed nine indicator variables corresponding to the first nine deciles of the DTD measure's unconditional distribution. We then interact each indicator variable with the securities market variables to generate nine sets of regression results. To focus the analysis, we use a more parsimonious model that has the lagged BOPEC rating as the only supervisory variable; i.e.,

$$\mathrm{BP}^*_{it} = \beta \mathrm{BP}_{it-2} + (\pi + \alpha_n \mathrm{I}_{nit-1})\mathrm{z}_{it-1} + \varepsilon_{it}, \tag{7}$$

where z_{it-1} is the securities market variable in question and I_{nit-1} is equal to one if BHC i's DTD measure is in the nth percentile at time $t-1$, and zero otherwise. Our main object of interest is whether the coefficients on the securities market variables are different in magnitude depending on how close the BHC is to its default point.

The estimation results for the 1,291 BOPEC ratings for BHCs with publicly traded equity are presented in table 5A. For the SCAR

Table 5A. Estimation Results for the Simplified Rating Model Using DTD Interactions with SCARs

DTD Pctile.	π for SCAR Variable	p-value	α_n Interaction with DTD Measure	p-value	$\pi + \alpha_n$	p-value	Pseudo R^2
10th Pctile.	−0.421	0.00	0.137	0.48	−0.284	0.14	0.31
20th Pctile.	−0.406	0.00	0.022	0.89	−0.384	0.01	0.31
30th Pctile.	−0.423	0.00	0.054	0.71	−0.369	0.00	0.31
40th Pctile.	−0.431	0.00	0.058	0.67	−0.373	0.00	0.31
50th Pctile.	−0.418	0.00	0.030	0.84	−0.388	0.00	0.31
60th Pctile.	−0.313	0.00	−0.114	0.42	−0.427	0.00	0.31
70th Pctile.	−0.271	0.04	−0.155	0.32	−0.426	0.00	0.31
80th Pctile.	−0.286	0.09	−0.126	0.49	−0.412	0.00	0.31
90th Pctile.	−0.197	0.41	−0.214	0.40	−0.411	0.00	0.31

Note: The coefficient estimates presented here are for the regression in equation (9). The securities market variable of interest is the BHC SCAR as used previously. The interacted indicator variable is equal to one if the BHC's distance-to-default (DTD) measure is in the nth percentile of its unconditional distribution. All models are estimated with a sample of 1,291 observations.

variable described previously, we expect its aggregate coefficient to be negative, suggesting that an increased SCAR that reflects improved BHC condition should lead to a better BOPEC rating. We expect the disaggregated π coefficient to be negative across all nine regressions. We expect the α_n coefficient to have its largest impact for the lowest DTD percentiles; i.e., the stock market signals for BHCs relatively closer to their default point should be less clear for supervisory purposes. As shown in the table, the estimated π coefficients are invariably negative and statistically significant for all but the ninetieth DTD percentile. The α_n coefficients are estimated to be positive for closer-to-default deciles and become negative as the default decile is increased, but they are not significantly different from zero. Taken as a whole, the equity market impact measured as the $\pi + \alpha_n$ values is negative, statistically significant for all but the tenth percentile specification, and in a relatively narrow range. Thus, as expected, equity market signals are weakest for BHCs relatively close to the default point.

In table 5B, the z_{it-1} variable is the adjusted BHC yield spread. The regressions are based on 283 BOPEC ratings assigned to BHCs

Table 5B. Estimation Results for the Simplified Rating Model Using DTD Interactions with the Adjusted Yield Spreads

DTD Pctile.	π for Adj. Yield Spread	p-value	α_n Interaction with DTD Measure	p-value	$\pi + \alpha_n$	p-value	Pseudo R^2
10th Pctile.	0.626	0.56	8.906	0.00	9.532	0.00	0.39
20th Pctile.	0.245	0.86	1.445	0.56	1.690	0.42	0.39
30th Pctile.	1.180	0.55	−0.651	0.77	0.529	0.61	0.39
40th Pctile.	1.956	0.22	−1.726	0.36	0.230	0.82	0.39
50th Pctile.	2.894	0.01	−3.099	0.05	−0.205	0.85	0.40
60th Pctile.	1.646	0.16	−1.095	0.54	0.551	0.66	0.39
70th Pctile.	1.859	0.21	−1.315	0.52	0.544	0.66	0.39
80th Pctile.	0.116	0.06	0.733	0.76	0.849	0.46	0.39
90th Pctile.	−1.631	0.50	2.499	0.33	0.868	0.43	0.39

Note: The coefficient estimates presented here are for the regression in equation (9). The securities market variable of interest is the BHC adjusted yield spread as used previously. The interacted indicator variable is equal to one if the BHC's distance-to-default (DTD) measure is in the nth percentile of its unconditional distribution. All models are estimated with a sample of 283 observations.

with both public equity and debt. For this variable, we expect the π coefficients to be positive, such that an increase in adjusted yield spreads suggests a worsening of BHC condition and probably a worsening of the BOPEC rating. We expect the α_n coefficients to be decreasing and diminishing the overall impact of the market signal as the DTD measure increases; i.e., the debt market signal should be strongest for BHCs closer to their default points. In the table, we show that, in fact, the coefficient of the estimated $(\pi + \alpha_1)$ coefficient for the percentile closest to the default point is more than four times larger than the coefficients on the other percentiles. Furthermore, that first coefficient is the only one that is statistically significant, due to the statistical significance of the α_1 coefficient. The α_n coefficients for the higher DTD percentiles are not statistically significant.

In summary, our simplified ordered logit models suggest an asymmetric contribution of equity and debt market signals to explaining BOPEC ratings, and the asymmetry depends on how close the BHC is to its default point. The impact of the equity market signals

$(\pi + \alpha_n)$ is largest in magnitude for the BHCs further from default, while the impact of the adjusted yield spreads is much larger in magnitude for BHCs very close to default.[17] These results are in line with theoretical expectations as well as prior empirical research. In particular, they provide further empirical insight into the finding by Berger, Davies, and Flannery (2000) that supervisory and debt-related assessments of BHCs are related. Furthermore, the results provide a refinement of their finding that supervisory and equity market assessments are not interrelated; that is, abnormal stock returns provide useful information with respect to changes in supervisory assessments only for BHCs further from default.

4. Out-of-Sample Forecast Performance

While the BOM model's in-sample fit and inference is interesting and important, a stricter test of its usefulness for supervisory purposes is its out-of-sample forecast accuracy. As in Krainer and Lopez (2003, 2004), we propose to estimate our model over rolling, four-quarter sample periods and evaluate their out-of-sample forecast performance. Clearly, the number of BOPEC ratings assigned with these rolling subsamples will be small relative to the full sample, but we accept the smaller estimation samples in order to simulate how securities market data might be used in practice by supervisors.[18]

However, a major difficulty in applying this forecasting approach to the four BOM models discussed in section 3 is that the number of observations in any given rolling subsample could be too small for the models with securities market data. That is, the number of BHCs with publicly traded securities within a given forecasting

[17]This result regarding adjusted BHC yield spreads seems to be consistent with Hanweck and Spellman (2002), who found that BHC subordinated debt yields were sensitive to solvency concerns if investors' forbearance expectations (i.e., time to closure or default) were less than one year.

[18]The need for four quarters of prior data also causes us to lose some of the early BOPEC assignments, mainly because we do not have bond market data prior to 1990. The sample used in the forecasting exercise consists of 2,878 BOPEC assignments. In this sample there were 638 upgrades, 331 downgrades, and 1,909 no-change inspections corresponding to 509, 290, and 866 unique BHCs, respectively.

window may be too small to permit model estimation. This is particularly true for the two samples of BHCs with publicly traded debt. To address this issue, we propose to pool the data and estimate a single model using a technique outlined in Griliches (1986). The underlying assumption to justify this pooling is that all ratings are generated by the same underlying process. In other words, for the equity sample, publicly traded BHCs are allowed to be different from private BHCs, but these differences are assumed to be observable. Thus, the relationship between a variable—say, the capital ratio—and the BOPEC rating is the same across publicly traded and private BHCs, holding the other covariates in the model fixed. If this assumption is satisfied, then we can treat the absence of a variable—say, stock market returns for BHCs that are not publicly traded—as though it were simply a case of missing data. We can then replace the missing values with the in-sample mean of the stock market return variable and estimate the model in equation (2) on the pooled sample, including fixed effects for the observations wherever we make this data adjustment. Formally, the estimation equation becomes

$$\mathrm{BP}^*_{it} = \beta \mathrm{x}_{it-2} + \pi \mathrm{z}_{it-1} + \mathrm{I}_{Eit-1} + \mathrm{I}_{Dit-1} + \varepsilon_{it}, \tag{8}$$

where the indicator variables I_E and I_D take on the value of 1 if BHC i has publicly traded equity and debt, respectively.

This estimation approach has two advantages. First, the coefficient estimates on the securities market variables should not be affected, since we are simply replacing the missing values with the variables' mean. Second, the pooling should allow us to get more efficient estimates of the coefficients on the supervisory variables because we can use all available observations.

To confirm that this procedure does not impact the empirical results, we conducted two robustness tests. First, we checked whether the distribution of BOPEC ratings for publicly traded BHCs differed from that of private BHCs. On average, we found no statistical differences, suggesting that the subsamples based on public equity and/or debt are not markedly different from the whole sample with respect to assigned BOPEC ratings. Second, we compared the in-sample estimation results in tables 4A and 4B with those corresponding to the appropriate specification of equation (8) using the Griliches adjustment. For example, for the case of the equity BOM model, we compare the in-sample estimation results based on the sample of 1,291 inspections of BHCs with publicly traded equity to the results where we use the full sample of inspections and plug in the mean abnormal return values for BHCs with no stock return data. Table 6 shows that the estimates for the securities market variables are very similar. The largest difference is observed for the sample based on BHCs with both public equity and debt, as expected, since it has the smallest sample size. However, even here, the coefficients' signs and significance are similar across the estimation methods. Given that the Griliches adjustment does not alter our results markedly and permits us to conduct our forecasting exercise for all four ratings samples, our forecasting results presented below are based on estimating equation (8) across our rolling data subsamples using the Griliches adjustment.[19]

4.1 Forecasting Results

Our measure of whether a model forecasts well is to ask how often its directional forecasts are correct. For example, if the model generates a signal suggesting a BOPEC upgrade, what percentage of the time does an upgrade actually take place? For our purposes, we generate a forecasted BOPEC rating $\widehat{\mathrm{BP}}_{it}$ as follows:

$$\widehat{\mathrm{BP}}_{it} = \sum_{j=1}^{5} \mathrm{j} * \widehat{\Pr}(\mathrm{BP}_{it} = \mathrm{j}), \tag{9}$$

where $\widehat{\Pr}(\mathrm{BP}_{it} = \mathrm{j})$ is the ordered logit model's forecasted probability based on information at time $t-1$ that BHC i will be assigned a rating of j at time t. We then compare these forecasted ratings to BP_{it-1}, the ratings as of time $t-1$. An upgrade (or downgrade) signal is generated when the forecasted rating is a full rating better (or worse) than the existing rating. That is, an upgrade signal is given when $\mathrm{BP}_{it-1} - \widehat{\mathrm{BP}}_{it} \geq 1$, and a downgrade signal is

[19]In addition, we performed a third robustness check by conducting the out-of-sample exercise for the equity BOM model based on both adjusted and unadjusted data. Comparisons of these two sets of forecasts are both qualitatively and quantitatively similar. The results from these robustness checks are available from the authors upon request.

Table 6. Coefficient Estimates on the Securities Market Variables for the Subsample and Griliches Adjustment Estimation Techniques

	Equity Only		Debt Only		Equity and Debt	
	Subsample Estimation	**Griliches Adjustment**	**Subsample Estimation**	**Griliches Adjustment**	**Subsample Estimation**	**Griliches Adjustment**
SCSR	−0.831*	−0.828*	–	–	−0.972*	−0.804*
SCAR	−0.523*	−0.529*	–	–	−0.552*	−0.522*
Adj. Yield Spread	–	–	+2.686*	+2.387*	+2.382	+2.234*
Sample Size	1,291	3,010	309	3,010	283	3,010

Note: A * represents significance at the 5 percent level.

given when $\mathrm{BP}_{it-1} - \widehat{\mathrm{BP}}_{it} \leq -1$. A "no change" signal occurs when $|\mathrm{BP}_{it-1} - \widehat{\mathrm{BP}}_{it}| < 1$. Note that we only generate one-quarter-ahead forecasts, even though we evaluate them at various horizons.

In tables 7A–D, we report the percentages of signals that are correct. As shown in table 7A for the core model, an upgrade signal received four quarters prior to a BOPEC assignment results in an actual upgrade 55 percent of the time; 36 percent of the time, the actual outcome of the inspection is no change in rating; and 9 percent of the time, the actual outcome is a downgrade. By one quarter prior to the inspection, the upgrade signal is accurate 90 percent of the time. The model appears to be just as effective at picking up downgrades. Four quarters prior to the inspection, a downgrade signal is accurate 68 percent of the time, improving to 91 percent accuracy one quarter prior to the inspection.

To evaluate the BOPEC forecasts and their directional signals, we compare the conditional probabilities in the tables to a benchmark. For the core BOM, a natural benchmark for comparison purposes is the unconditional distribution of BOPEC ratings changes. In our sample, the unconditional probabilities of upgrades, downgrades, and no changes are 22 percent, 12 percent, and 66 percent, respectively. That is, given no information about a BHC, the probability of an upgrade is 22 percent. Given a signal from the core BOM, the upgrade probability rises to 55 percent at the four-quarter horizon and to 90 percent at the one-quarter horizon; see table 7A. Thus, conditioning on supervisory information in the core BOM model is clearly useful relative to the unconditional probabilities. This notion is formalized by the Pearson goodness-of-fit results in the rightmost column, which tests whether the model's conditional change probabilities are statistically different from the unconditional probabilities. We easily reject the null hypothesis that the conditioning variables do not provide useful information for forecasting BOPEC ratings.

As shown in tables 7B through 7D, forecast performance does not change much when securities market data are added into the BOM model; in fact, performance is a bit worse in some cases. With respect to the equity BOM model, the results in table 7B show that upgrade-signal accuracy increases from 64 percent four quarters prior to inspection to 91 percent accuracy within one quarter of the inspection, compared to 55 percent and 90 percent, respectively,

Table 7A. Out-of-Sample Forecast Performance of the Core BOM Model

	# Signals	Actual Inspection Outcome			Pearson Statistic
		Upgrade %	No Change %	Downgrade %	
Signal at −4 Quarters					
Upgrade	22	**55%**	36%	9%	
No change	2,825	22%	**67%**	11%	
Downgrade	31	3%	29%	**68%**	145.0*
Signal at −3 Quarters					
Upgrade	28	**68%**	21%	11%	
No change	2,820	22%	**67%**	11%	
Downgrade	30	0%	17%	**83%**	254.3*
Signal at −2 Quarters					
Upgrade	45	**80%**	13%	7%	
No change	2,793	22%	**68%**	10%	
Downgrade	40	0%	10%	**90%**	460.1*
Signal at −1 Quarter					
Upgrade	60	**90%**	8%	2%	
No Change	2,773	21%	**69%**	10%	
Downgrade	45	0%	9%	**91%**	624.6*

Note: This table presents the forecast performance results based on the 2,878 BOPEC change signals from the core BOM model at different horizons. A forecast signal is the difference between the forecasted BOPEC rating and the previously assigned BOPEC rating. Signals greater than one (less than one) are forecasts of upgrades (downgrades), respectively. The figures in bold indicate the outcome expected, conditional on the signal. Percentages in rows may not sum to 100 percent due to rounding. The Pearson goodness-of-fit statistic tests the null hypothesis that the distribution of BOPEC change outcomes conditional on the core model forecasts is not different from the unconditional probabilities of BOPEC upgrades (22 percent), no changes (66 percent), and downgrades (12 percent). The statistic is distributed $\chi^2(10)$. A * denotes significance at the 5 percent level.

Table 7B. Out-of-Sample Forecast Performance of the Equity BOM Model

	# Obs.	Actual Inspection Outcome			Pearson Stat. I	Pearson Stat. II
		Upgrade %	No Change	Downgrade %		
Signal at −4 Quarters						
Upgrade	25	**64%**	28%	8%		
No Change	2,815	22%	**67%**	11%		
Downgrade	38	0%	45%	**55%**	131.8*	12.4*
Signal at −3 Quarters						
Upgrade	30	**73%**	20%	7%		
No Change	2,810	22%	**68%**	11%		
Downgrade	38	0%	16%	**84%**	326.2*	1.2
Signal at −2 Quarters						
Upgrade	49	**84%**	12%	4%		
No Change	2,771	22%	**68%**	10%		
Downgrade	58	0%	10%	**90%**	627.1*	1.5
Signal at −1 Quarter						
Upgrade	69	**91%**	7%	1%		
No Change	2,744	21%	**69%**	10%		
Downgrade	65	0%	17%	**83%**	755.0*	7.7*

Note: This table presents the forecast performance results based on the 2,878 BOPEC change signals from the equity BOM model at different horizons. A signal is the difference between the forecasted BOPEC rating and the previously assigned BOPEC rating. Signals greater than one (less than one) are forecasts of upgrades (downgrades), respectively. The figures in bold indicate the outcome expected, conditional on the signal. Percentages in rows may not sum to 100 percent due to rounding. The Pearson goodness-of-fit statistic I tests the null hypothesis that the distribution of BOPEC change outcomes conditional on the equity BOM model forecasts is not different from the in-sample probabilities of BOPEC upgrades (22 percent), no changes (66 percent), and downgrades (12 percent). The statistic is distributed $\chi^2(10)$. The Pearson goodness-of-fit statistic II tests the null hypothesis that the distribution of BOPEC change outcomes conditional on the equity BOM model forecasts is not different from the distribution of outcomes forecasted by the core BOM model. The statistic is distributed $\chi^2(2)$. A * denotes significance at the 5 percent level.

Table 7C. Out-of-Sample Forecast Performance of the Debt BOM Model

	# Obs.	Actual Inspection Outcome: Upgrade %	No Change	Downgrade %	Pearson Stat. I	Pearson Stat. II
Signal at −4 Quarters						
Upgrade	25	**52%**	40%	8%		
No Change	2,818	22%	**67%**	11%		
Downgrade	35	3%	37%	**60%**	123.1*	3.5
Signal at −3 Quarters						
Upgrade	27	**67%**	22%	11%		
No Change	2,818	22%	**67%**	11%		
Downgrade	33	0%	21%	**79%**	246.0*	0.9
Signal at −2 Quarters						
Upgrade	43	**81%**	12%	7%		
No Change	2,792	22%	**68%**	10%		
Downgrade	43	0%	12%	**88%**	470.8*	1.5
Signal at −1 Quarter						
Upgrade	60	**88%**	8%	3%		
No Change	2,765	21%	**69%**	10%		
Downgrade	53	0%	11%	**89%**	657.2*	1.9

Note: This table presents the forecast performance results based on the 2,878 BOPEC change signals from the debt BOM model at different horizons. A forecast signal is the difference between the forecasted BOPEC rating and the previously assigned BOPEC rating. Signals greater than one (less than one) are forecasts of upgrades (downgrades), respectively. The figures in bold indicate the outcome expected, conditional on the signal. Percentages in rows may not sum to 100 percent due to rounding. The Pearson goodness-of-fit statistic I tests the null hypothesis that the distribution of BOPEC change outcomes conditional on the debt BOM model forecasts is not different from the in-sample probabilities of BOPEC upgrades (22 percent), no changes (66 percent), and downgrades (12 percent). The statistic is distributed $\chi^2(10)$. The Pearson goodness-of-fit statistic II tests the null hypothesis that the distribution of BOPEC change outcomes conditional on the debt BOM model forecasts is not different from the distribution of outcomes forecasted by the core BOM model. The statistic is distributed $\chi^2(2)$. A * denotes significance at the 5 percent level.

Table 7D. Out-of-Sample Forecast Performance of the Extended BOM Model

	# Obs.	Actual Inspection Outcome			Pearson Stat. I	Pearson Stat. II
		Upgrade %	No Change	Downgrade %		
Signal at −4 Quarters						
Upgrade	28	**57%**	36%	7%		
No Change	2,813	22%	**67%**	11%		
Downgrade	37	0%	43%	**57%**	126.7*	6.4*
Signal at −3 Quarters						
Upgrade	31	**68%**	267%	6%		
No Change	2,808	22%	**67%**	11%		
Downgrade	39	0%	23%	**77%**	270.5*	3.3
Signal at −2 Quarters						
Upgrade	48	**83%**	13%	4%		
No Change	2,767	22%	**68%**	10%		
Downgrade	63	0%	13%	**87%**	633.1*	3.1
Signal at −1 Quarter						
Upgrade	67	**90%**	7%	3%		
No Change	2,746	21%	**69%**	10%		
Downgrade	65	2%	17%	**83%**	715.5*	11.9*

Note: This table presents the forecast performance results based on the 2,878 BOPEC change signals from the extended BOM model at different horizons. A forecast signal is the difference between the forecasted BOPEC rating and the previously assigned BOPEC rating. Signals greater than one (less than one) are forecasts of upgrades (downgrades), respectively. The figures in bold indicate the outcome expected, conditional on the signal. Percentages in rows may not sum to 100 percent due to rounding. The Pearson goodness-of-fit statistic I tests the null hypothesis that the distribution of BOPEC change outcomes conditional on the extended BOM model forecasts is not different from the in-sample probabilities of BOPEC upgrades (22 percent), no changes (66 percent), and downgrades (12 percent). The statistic is distributed $\chi^2(10)$. The Pearson goodness-of-fit statistic II tests the null hypothesis that the distribution of BOPEC change outcomes conditional on the extended BOM model forecasts is not different from the distribution of outcomes forecasted by the core BOM model. The statistic is distributed $\chi^2(2)$. A * denotes significance at the 5 percent level.

for the core model. For downgrades, however, forecast performance improves from 55 percent to 83 percent as the inspection approaches, which is not quite as large an improvement as for the core model. Overall, the Pearson test results in the last column suggest that the addition of the equity market variables does not improve forecast performance relative to the core model. In fact, at four quarters and one quarter prior, these variables reduce forecast performance.

In table 7C, the debt BOM model's forecasting performance is actually a bit worse than that of the core model with only supervisory variables. For example, an upgrade signal four quarters prior to the inspection is correct 52 percent of the time, while a downgrade signal at four quarters prior results in an actual downgrade 60 percent of the time. When compared to the core model's forecast signals using the Pearson test, we cannot reject the null hypothesis that the debt market variables do not improve forecast performance. Finally, as shown in table 7D, the forecast performance of the extended model with both equity and debt market variables is again marginally worse than that of the core model. In summary, the introduction of securities market variables to the core BOM model fails to improve overall BOPEC forecast performance for rating changes at conventional levels of significance.

4.2 Information in the Forecasts

Although BOPEC change forecasts do not appear to be appreciably different across the core and the extended BOM models presented in tables 7A–D, the set of BOPEC rating changes correctly signaled by these models are not identical. This outcome suggests the need to think carefully about the cost of forecast errors to supervisors. The forecasting literature has shown that combining forecasts from different models can improve certain aspects of forecast performance; see Granger and Newbold (1986) as well as Diebold and Lopez (1996) for further discussion. Hence, another way to gauge the contribution of securities market information is to examine the additional forecast signals for BHCs with public securities as generated by the extended models relative to the core model's signals. Seen in this light, the marginal benefit of monitoring securities market information is notable.

Table 8A. Improvements in BOPEC Downgrade Signals

	Equity BOM Model	Debt BOM Model	Extended BOM Model
4 Quarters Prior	24%	14%	22%
3 Quarters Prior	28%	10%	31%
2 Quarters Prior	28%	9%	29%
1 Quarter Prior	24%	5%	25%

Note: This table presents the percentage improvement in correct BOPEC downgrade signals when combining the downgrade signals from the core BOM model with those from the models incorporating securities market data. Downgrade signal is defined as forecasted rating − current rating > 1. The table reports the number of downgrades correctly signaled by the alternative models and not identified by the core model, expressed as a percentage of downgrades correctly identified by the core model. Sample contains 2,878 inspections, with 331 downgrades.

Comparing BOPEC change signals generated by the core and the three extended BOM models, we ask what is the percentage increase in the number of correct downgrade signals when securities market data are incorporated into the core BOM model?[20] We focus on downgrades because these are the events of most interest to supervisors. As reported in table 8A, the extended model with adjusted bond yields and stock return data produces an additional 22 percent more correct signals at the four-quarter horizon over and above those produced by the core model. At the one-quarter horizon, the improvement is 25 percent more correct signals. Another interesting point is the similarity between the marginal contributions of the equity BOM model and the extended BOM model with both equity and debt market variables. Evidently, in this particular framework, most of the additional downgrade signals a supervisor can extract from securities market data come from the equity market. This result contrasts with the in-sample results and may be due to the relatively small number of BHCs with publicly traded debt in any given subsample period.

Of course, these three BOM models also produce incorrect signals over and above those produced by the core model. Since table 8A

[20]Note that all of these additional correct forecasts are for ratings changes at BHCs with publicly traded securities.

Table 8B. Trade-Off between Correct and Incorrect Downgrade Signals

	Equity BOM Model	Debt BOM Model	Extended BOM Model
4 Quarters Prior	1/2	7/11	11/25
3 Quarters Prior	17/12	6/7	19/17
2 Quarters Prior	23/17	7/4	6/5
1 Quarter Prior	3/2	5/1	5/3

Note: This table presents the trade-off between additional correct and additional incorrect BOPEC downgrade signals provided by the alternative BOM models relative to the core BOM model. A cell entry of x/y suggests that the alternative BOM model identifies x additional correct downgrades signals beyond those of the core model, at the rate of y additional incorrect downgrade signals. Sample contains 2,878 inspections, with 331 downgrades.

shows that these models identify additional BOPEC downgrades, their mistakes may be responsible for our earlier result that their overall forecast performance was almost the same as the core model. We examine this trade-off more closely in table 8B, where we express the models' ratios of correct downgrade signals to incorrect signals, which are also known as false positives or type 2 errors. For the extended BOM model at the four-quarter horizon, the model produces eleven additional correct signals at the cost of twenty-five incorrect downgrade signals. By the one-quarter horizon, however, the performance improves dramatically to five additional correct signals at the cost of only three additional incorrect signals. The models extended by equity variables and debt variables alone behave similarly.[21] Interestingly, this signal trade-off for the debt BOM model is quite good; by one quarter out, it produces five correct signals for every incorrect signal. However, as indicated in table 8A, the

[21]Gropp, Vesala, and Vulpes (2004, 2006) conduct a similar analysis, but they focus on in-sample model fit. In their analysis, incorporating securities market variables into their model reduces false positive errors (i.e., type 2 errors). Our results align well with this result in that the combination of BOPEC downgrade signals from the core BOM model with the three expanded models adds more correct than incorrect signals, especially nearer to the BOPEC assignment; i.e., the combined set of signals has a reduced type 2 error rate than the set based just on the core model.

drawback to this model is that it produces relatively fewer signals beyond those from the core model.

In summary, our analysis of combined BOPEC downgrade signals indicates that models using securities market data can generate a reasonably large number of additional correct downgrade signals relative to the core BOM model. Given the emphasis placed by supervisors on BOPEC downgrades, we believe that these forecasting results argue for closer monitoring of securities market information for off-site supervisory monitoring.

5. Conclusion

Our empirical results indicate that both equity and debt market information are useful in improving the in-sample fit of our proposed BOM model for BOPEC ratings. Both types of securities market information also appear to be useful in explaining BOPEC upgrades and downgrades. Moreover, for our data set, we detect nonlinearities in the impact of securities market variables on BOPEC ratings. For BHCs closer to their estimated default points, the effect of our adjusted BHC bond yield spreads on BOPEC ratings is larger in magnitude than it is for BHCs further from their default points, and vice versa for equity market data. These results provide further empirical support for the finding by Berger, Davies, and Flannery (2000) that supervisory and debt-related assessments of BHCs are correlated and a refinement of their finding that supervisory and equity market assessments are not related; i.e., abnormal BHC stock returns provide useful information regarding supervisory assessments for BHCs that are far from their default points.

When we turn to out-of-sample forecasting, however, evidence for the usefulness of market information is disappointingly weak. For our analysis, we estimate our four BOM model specifications on a rolling subsample of data and then forecast BOPEC ratings out of sample. We find the forecast performance of the three BOM models that include securities market data is not much different than the performance of the core model based on supervisory data alone.

While the overall forecasting performance of the four BOM models is similar, the sets of forecasted rating changes are not identical. That is, the core model correctly identifies one set of BOPEC rating changes, particularly downgrades, while the extended models correctly identify other sets of downgrades. The extended BOM models correctly identify additional BOPEC downgrades for publicly traded BHCs over and above the correct forecasts from the core model. These additional correct forecasts can be achieved at a relatively modest cost of additional incorrect signals. Hence, supervisory use of securities market information within the context of an off-site monitoring model, such as our proposed BOM model, appears to be reasonable.

References

Basel Committee on Banking Supervision. 2003. "Markets for Bank Subordinated Debt and Equity in Basel Committee Member Countries." Working Paper No. 12, Bank for International Settlements.

Berger, A. N., S. M. Davies, and M. J. Flannery. 2000. "Comparing Market and Supervisory Assessments of Bank Performance: Who Knows What When?" *Journal of Money, Credit, and Banking* 32 (3): 641–67.

Bliss, R., and M. Flannery. 2001. "Market Discipline in the Governance of U.S. Bank Holding Companies: Monitoring versus Influencing." In *Prudential Supervision: What Works and What Doesn't*, ed. F. S. Mishkin, 107–43. Chicago: University of Chicago Press.

Board of Governors of the Federal Reserve System. 1999. "Using Subordinated Debt as an Instrument of Market Discipline." Report of the Study Group on Subordinated Notes and Debentures. Federal Reserve System Staff Study No. 172.

———. 2004. "Bank Holding Company Rating System." Supervision and Regulation Letter No. 04-18, Division of Banking Supervision and Regulation.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

Cole, R. A., and J. W. Gunther. 1998. "Predicting Bank Failures: A Comparison of On- and Off-Site Monitoring Systems." *Journal of Financial Services Research* 13 (2): 103–17.

Covitz, D. M., D. Hancock, and M. L. Kwast. 2004a. "A Reconsideration of the Risk Sensitivity of U.S. Banking Organization Subordinated Debt Spreads: A Sample Selection Approach." *Economic Policy Review* (Federal Reserve Bank of New York) 10 (2): 73–92.

———. 2004b. "Market Discipline in Banking Reconsidered: The Roles Funding Manager Decisions and Deposit Insurance Reform." Finance and Economics Discussion Series No. 2004-53, Board of Governors of the Federal Reserve System.

DeFerrari, L. M., and D. E. Palmer. 2001. "Supervision of Large Complex Banking Organizations." *Federal Reserve Bulletin* (February): 47–57.

DeYoung, R., M. J. Flannery, W. W. Lang, and S. M. Sorescu. 2001. "The Information Content of Bank Exam Ratings and Subordinated Debt Prices." *Journal of Money, Credit, and Banking* 33 (4): 900–25.

Diebold, F. X., and J. A. Lopez. 1996. "Forecast Evaluation and Combination." In *Handbook of Statistics 14: Statistical Methods in Finance*, ed. G. S. Maddala and C. R. Rao, 241–68. Amsterdam: North-Holland.

Estrella, A., S. Park, and S. Peristiani. 2000. "Capital Ratios as Predictors of Bank Failure." *Economic Policy Review* (Federal Reserve Bank of New York) 6 (2): 33–52.

Evanoff, D., and L. Wall, L. 2001. "Sub-debt Yield Spreads as Bank Risk Measures." *Journal of Financial Services Research* 20 (2): 121–46.

———. 2002. "Measures of the Riskiness of Banking Organizations: Subordinated Debt Yields, Risk-Based Capital and Examination Ratings." *Journal of Banking and Finance* 26 (5): 989–1009.

Flannery, M. J. 1998. "Using Market Information in Prudential Bank Supervision: A Review of the U.S. Empirical Evidence." *Journal of Money, Credit, and Banking* 30 (3): 273–305.

Flannery, M. J., and S. U. Sorescu. 1996. "Evidence of Bank Market Discipline in Subordinated Debenture Yields: 1983–1991." *Journal of Finance* 51 (4): 1347–77.

Goyal, V. K. 2005. "Market Discipline of Bank Risk: Evidence from Subordinated Debt Contracts." *Journal of Financial Intermediation* 14 (3): 318–50.

Granger, C. W. J., and Newbold, P. 1986. *Forecasting Economic Time Series*. 2nd ed. San Diego: Academic Press.

Griliches, Z. 1986. "Economic Data Issues." In *Handbook of Econometrics*, Vol. III, ed. Z. Griliches and M. D. Intriligator. Amsterdam: Elsevier Science Publishers BV.

Gropp, R., and A. J. Richards. 2001. "Rating Agency Actions and the Pricing of Debt and Equity of European Banks: What Can We Infer About Private Sector Monitoring of Bank Soundness?" *Economic Notes* 30 (3): 373–98.

Gropp, R., J. Vesala, and G. Vulpes. 2004. "Market Indicators, Bank Fragility, and Indirect Market Discipline." *Economic Review* (Federal Reserve Bank of New York) 10 (2): 53–62.

———. 2006. "Equity and Bond Market Signals as Leading Indicators of Bank Fragility." *Journal of Money, Credit, and Banking*, 38 (2): 399–428.

Gunther, J. W., and R. R. Moore. 2003. "Early Warning Models in Real Time." *Journal of Banking and Finance* 27 (10): 1979–2002.

Hancock, D., and M. Kwast. 2001. "Using Subordinated Debt to Monitor Bank Holding Companies: Is It Feasible?" *Journal of Financial Services Research* 20 (2): 147–87.

Hanweck, G. A., and L. J. Spellman. 2002. "Forbearance Expectations and the Subordinated Debt Signal of Bank Insolvency." Manuscript, School of Management, George Mason University.

Hirtle, B. J., and J. A. Lopez. 1999. "Supervisory Information and the Frequency of Bank Examinations." *Economic Policy Review* (Federal Reserve Bank of New York) 5 (1): 1–20.

Krainer, J., and J. A. Lopez. 2003. "How Might Financial Market Information Be Used for Supervisory Purposes?" *Economic Review* (Federal Reserve Bank of San Francisco), 29–45.

———. 2004. "Incorporating Equity Market Information into Supervisory Monitoring Models." *Journal of Money, Credit, and Banking* 36 (6): 1043–67.

Meyer, P. A., and H. W. Pifer. 1970. "Prediction of Bank Failures." *Journal of Finance* 25 (4): 853–68.

Pettway, R. H., and J. F. Sinkey. 1980. "Establishing On-Site Bank Examination Priorities: An Early-Warning System Using Accounting and Market Information." *Journal of Finance* 35 (1): 137–50.

Ronn, E. I., and A. K. Verma. 1986. "Pricing Risk-Adjusted Deposit Insurance: An Option-Based Model." *Journal of Finance* 41 (4): 871–95.

Sahajwala, R., and van den Bergh, P. 2000. "Supervisory Risk Assessment and Early Warning Systems." Basel Committee on Banking Supervision Working Paper No. 4, Bank for International Settlements.

Sinkey, J. F. 1975. "A Multivariate Statistical Analysis of the Characteristics of Problem Banks." *Journal of Finance* 30 (1): 21–36.