EFA
Eastern
Finance
Association

The Financial Review 37 (2001) 75–96

The Financial Review FR

Combining Bond Rating Forecasts Using Logit

Mark Kamstra*
Peter Kennedy
Teck-Kin Suan

Simon Fraser University

Abstract

Companies sometimes use statistical analysis to anticipate their bond ratings or a change in the rating. However, different statistical models can yield different ratings forecasts, and there is no clear rule for which model is preferable. We use several forecasting methods to predict bond ratings in the transportation and industrial sectors listed by Moody's bond rating service. A variant of the ordered-logit regression-combining method of Kamstra and Kennedy 1998 yields statistically significant, quantitatively meaningful improvements over its competitors, with very little computational cost.

Keywords: bond rating, forecast combining

JEL classifications: G10/G31/C52/C53

1. Introduction

Consensus forecasts, formed by combining competing forecasts, are popular. The simplest consensus forecast is the average of competing forecasts. Recently, Kamstra and Kennedy (1998) have proposed a combining technique that is tailored to the special features of forecasting qualitative variables. In this paper, we investigate the usefulness of Kamstra and Kennedy's logit-combining method for forecasting bond ratings. The two data sets used in this paper are new U.S. transportation bonds

*Corresponding author: Department of Economics, Simon Fraser University, Burnaby BC, Canada, V5A 1S6; Phone: (604) 291-4514; Fax: (604) 291-5944. E-Mail: kamstra@sfu.ca

We thank Lisa Kramer and R. Glen Donaldson for helpful comments and the Social Sciences and Humanities Research Council of Canada for financial support.

issued between 1989 and 1992, and new U.S. industrial bonds issued in 1993. Consistent with the combining literature, our empirical results show that in general combined forecasts outperform the forecasts being combined. We also find that a modification of the Kamstra and Kennedy method is the combining method of choice in this context, in particular that it outperforms the traditional method of averaging probability forecasts. A caveat is that the Kamstra and Kennedy method, although computationally straightforward, requires data sufficient to produce reliable estimates of weighting factors.

The paper is as follows. Section 2 reviews the bond rating and forecast-combining literatures and related issues. Section 3 discusses the data and Section 4 presents our classification methods. Section 5 examines issues related to combining ordinal forecasts. Section 6 presents the rules for evaluating forecast performance, and Section 7 discusses the estimation results. Section 8 concludes.

2. Overview

Corporate bond ratings are intended to provide investors with a guide to the relative investment qualities of bond issues. Firms spend a great deal of money to get their bond issues rated, and the rating of an issue is influential in determining the firm's credit risk and the marketability of its bonds. The size of the corporate bond market alone underscores the importance of anticipating the rating an agency is likely to give a bond issue. During the first three quarters of 1999, a total of $560.7 billion was issued in the U.S. corporate new bond market. The Bond Market Association of the United States of America estimated daily trading volume at $10 billion. In 1999, Moody's corporate bond yield averages ranged from 7% for Aaa to 7.9% for Baa bonds.

In the area of bond ratings research, the traditional approach in classifying and predicting bond ratings has been to use individual classification models, such as ordinary least squares (OLS), multinomial discriminant analysis, ordered/unordered logistic regression, and ordered probit regression. Examples are presented by Pogue and Soldofsky (1969), Pinches and Mingo (1973), Kaplan and Urwitz (1979), Belkaoui (1980), Fox (1984), Ederington (1985), Terza (1985), Gentry, Whitford and Newbold (1988), Kao and Wu (1990), and Moon and Stotsky (1993a,1993b).

Related studies include those of Hand, Holthausen, and Leftwich (1992), who find excess bond return effects accompanying rating changes; Impson, Karafiath, and Glasscock (1992) who look at the impact of bond re-grading on the firm's beta; Goh and Ederington (1993) who determine when to expect a stock return decline in reaction to a downgrade in bond rating; Datta, Iskandar-Datta, and Patel (1997) who find (among other things) that bond ratings are a significant determinant of bond IPO returns; Followill and Martell (1997) who look at the effect of bond rating reviews and ratings changes on equity returns; and O'Neal (1998) who finds that the bond rating of electric utilities has a strong impact on equity sensitivity to interest rates.

Belkaoui (1980) argues that predicting bond ratings can provide vital information on how rating agencies perform and on how a yet-unannounced bond issue would be rated. Other literature studies applications for bond-rating forecast models.

There is general consensus on the inappropriateness of least squares methods to rate bonds. There is also concern over the use of methods that ignore the ordinal nature of bond rating, such as multinomial discriminant analysis. However, there is no real guidance as to what sort of statistical model is appropriate for bond rating. Furthermore, statistical models explain only about two thirds of bond ratings. Some researchers, including Reiter and Ziebart (1991), suspect that the prediction models themselves might be misspecified. Although theory provides some guidance in the choice of explanatory variables, the choice of method (parametric versus non-parametric, logit versus probit) is not indicated by either theory or empirical evidence. Certainly, no single method dominates in the empirical literature. This ambiguity has led researchers in other fields to argue for consensus forecasting.

2.1. The forecast-combining literature

Forming consensus out of conflicting forecasts, or forecast combination, is an accepted practice in the forecasting literature. Clemen (1989) is a survey. Clemen's (1989) bibliography includes over 2000 journal pages and 11 books, monographs, and theses.

Clemen (1989), Hallman and Kamstra (1989), Gunter and Aksu (1989), Donaldson and Kamstra (1996,1999), and many other researchers consider Bates and Granger (1969) the seminal work in combining forecasts. Bates and Granger combine the individual forecasts by minimizing the variance of the combined error forecast errors. This method is also discussed in Newbold and Granger (1974) and Winkler and Makridakis (1983). Granger and Ramanathan (1984) show that the Bates and Granger method is essentially an OLS regression on individual forecasts, with the coefficients restricted to sum to unity and the intercept to zero. There is a similar discussion in Bopp (1985), Clemen (1986), and Trenkler and Linski (1986).

Granger and Ramanathan (1984) also demonstrate that the regression method with unconstrained coefficients yields combined forecasts with smaller mean squared errors (MSE) than does the Bates and Granger (1969) weighting method. Other forecast combination methods include Bunn's (1975) out-performance method, the Bayesian techniques (using posterior probabilities) discussed in Clemen and Winkler (1986), Bunn (1981), and Bordley (1986), and genetic algorithms demonstrated in Harrald and Kamstra (1997).

In a survey of forecasting methods and accuracy, Mahmoud (1984) concludes that combining forecasts can improve accuracy. Granger (1989) summarizes the usefulness of combining forecasts: "(T)he combination of forecasts is a simple, pragmatic, and sensible way to possibly produce better forecasts." Clemen (1989) observes that combining forecasts increases forecasting accuracy, whether the forecasts are subjective, statistical, econometric, or by extrapolation, and that averaging forecasts in many cases dramatically improves forecast performance.

In Winkler and Makridakis (1983) averaging outperforms individual forecasts and accuracy improves with additional forecasts, but the improvement levels off eventually. Makridakis and Winkler (1983) use five different weighting procedures and find that two methods are more accurate than either the individual forecasts or the simple unweighted average.

All these combining methods deal primarily with quantitative variables. These methods should not be applied directly to bond ratings studies, such as this one, in which the dependent variable is ordinal and polychotomous, because such discrete dependent variables present distinct estimation problems. We mention them here because these methods are what the terminology "combining forecasts" brings to mind. We wish to emphasize that the novelty of our contribution in this paper stems from employing a combining technique that is better suited to this context.

Forecast combination work dealing with discrete dependent variable forecasts can be broadly classified into two types, one that combines ordinal forecasts directly, and one that combines probability forecasts. The literature that investigates schemes to combine ordinal forecasts directly is small, including only Feather and Kaylen (1989), and Fan, Lau, and Leung (1996), both of which look at binary forecasting situations. A larger literature examines combining the probability forecasts of a binary event. Sanders (1963) averages the probability forecasts from individual forecasters and concludes that average forecasts perform better than individual forecasts. Staël von Holstein (1971, 1972) and Winkler, Murphy, and Katz (1977) conduct similar experiments and come to similar conclusions. They also find that the average performs as well as or better than more complicated updating-of-weights schemes. Clemen's (1989) survey article provides an extensive review of this literature.

Recent work of Kamstra and Kennedy (1998) integrates these two approaches with logit-based forecast-combining methods that are applicable to dichotomous, polychotomous, and ordered-polychotomous contexts. In their methods, the forecasts to be combined can be probability forecasts, themselves ordinal (qualitative), or a mixture of the two. The Kamstra and Kennedy (1998) method most closely follows the convention of the literature on combining quantitative forecasts. It allows researchers to estimate weights for alternative forecasts in a computationally convenient manner (via logit, multinomial-logit, or ordered-logit regression) and to do so in a fashion that can correct for bias. Unlike earlier work on combining probability forecasts, Kamstra and Kennedy find that if there is sufficient variation among the forecasts being combined and an adequate number of observations, the logit-combining method can outperform the simple average.

2.2. The bond-rating analysis literature

Previous studies show that bond ratings can be predicted by a small number of explanatory variables. In general, these variables measure leverage, profitability/

Table 1

Summary of previous work

We report summary statistics for previous studies of forecasting bond ratings. For each study we report both the percentage of bonds that were rated correctly in the estimation sample and in the hold out or out of sample period, for every forecasting method used in the study.

	Model	Estimation Sample % Predicted Correctly	Hold out Sample % Predicted Correctly
Ederington (1985)	linear regression	56.5	65.0
	ordered probit regression	65.0	78.0
	multinomial discriminant analysis (proportional priors)	59.8	69.0
	unordered logit	70.3	73.0
	multinomial discriminant analysis (equal priors)	55.7	64.0
	quadratic discrimination analysis	63.0	72.0
Kaplan and Urwitz (1979)	linear regression		55.0
	ordered probit regression		50.0
Gentry, Whitford and Newbold (1988)	ordered probit regression		41.0
Belkaoui (1980)	multinomial discriminant analysis	62.5	65.9
Terza (1985)	ordered probit regression		67.0

interest coverage, firm size, and subordination status. For instance, Ederington (1985) uses subordination, total assets, long-term debt to capital ratio, and interest coverage as explanatory variables. Kaplan and Urwitz (1979) use interest coverage, long-term debt to total assets, long-term debt to net worth, net income to total assets, total assets, size of bond issue, coefficient of variation of total assets, coefficient of variation of net income, and subordination status. Gentry, Whitford, and Newbold (1988) and Fox (1984) use subordination, size of issue, debt ratio, cumulative years that dividends were paid, net income to total assets, and net income to interest.

Gentry, Whitford, and Newbold (1988) find that combining 12 funds' flow components with six financial ratios does not improve the classification ability of the ordered probit regression model. The authors' success in predicting bond ratings varies from 41% to 78% of the sample. Table 1 shows estimation results of a sample of previous studies on bond-rating prediction.

In analyzing municipal bond ratings, Moon and Stotsky (1993a, 1993b) note that municipalities choose to have their bonds rated, and will be more likely to do so if they have reason to believe that their bonds will be rated favorably. This phenomenon gives rise to sample selection bias in statistical estimation of the factors affecting bond ratings. In our paper, companies are trying to forecast their bond ratings. To do so they will use available data, such as the data we employ in this paper. These data could be characterized by selection bias. To improve forecasts, companies could employ a sophisticated correction procedure such as that suggested by Moon and Stotsky. But they may choose not to do so because of the complexity of this correction procedure or because data required for this procedure are unavailable. Regardless of what choice they make in this regard, it is unlikely that the results of this paper, concerning the appropriate way to combine forecasts, will be affected. Accordingly, in this paper we confine our attention to the case in which companies do not attempt to correct for selection bias.

The actual rating process involves more than just a handful of quantifiable statistical variables. The rater takes into account other unquantifiable variables such as leadership quality, management ability, and technology changes. It is not surprising that quantitative methods can only correctly predict about 78% of the observed ratings in these studies. Pogue and Soldofsky (1969) quote the following from an industry executive:

> *"... bond ratings, while based on statistics, are not entirely based on statistics. When the professional rater has studied all the facts, he is still left with the intangibles of judgment which involves personalities, technological possibilities, and just plain hunches."

If statistical analysis cannot explain bond ratings well, why would we be interested in predicting bond ratings? First, our predictions can be used as a check of a rater's assessments. Any large discrepancy should alert the rater to the possibility of errors. Second, getting a bond rating is expensive. A company might decide to use these statistical methods to estimate what rating it can expect before it goes to the rating agencies, or use these methods to anticipate a change in its rating. Companies are motivated to perform this sort of preliminary analysis because bond ratings affect financing costs and other determinants of profitability.

3. Data

We use two data sets in this paper. The Transport data set comprises new transportation bonds issued between 1989 to 1992 in the U.S. The Industrial data set contains new U.S. industrial bonds issued in 1993. We use the Moody's Investors Service, Inc. manuals (1988-1993) as our information source.

After we discard listings with inadequate information, we find that Moody's lists 89 new bond issues for Transport. The transport sector includes airlines, railroads, couriers, car rental agencies, and port authorities.

Industrial has 265 observations of newly issued industrial bonds in 1993. Companies listed in Moody's include manufacturers, wholesalers, retailers, hotels, and service companies.

The rating categories for Transport and for Industrial range from Aaa to B. For model estimations, we merge the Transport data set category B with Ba because there was only one B category observation. Table 2 details the number of observations in each category for the two data sets.

We use independent variables similar to those used in previous studies, most notably Ederington 1985. Table 3 defines the variables used in our bond ratings

Table 2

Observations per bond rating

We report the number of observations per bond rating category for our two data sets.

B	1	62
Ba	8	36
Baa	45	75
A	17	65
Aa	12	22
Aaa	6	5
Total	89	265

Table 3

Variables used in the bond ratings prediction

We report descriptions of the variables used as explanatory variables in our forecasting models.

Variable Description/Definition
Interest coverage: net income plus interest expense, divided by interest expense
Debt ratio: total debt divided by total assets
Return on assets: net income divided by total assets
Total firm assets: book value of firm's assets
Subordination status: 1 if the debt issue has seniority, and 0 otherwise

study and Table 4 lists summary statistics for the data. Because we collect the Transport data over several years, we adjust the monetary variable ''total firm assets'' for inflation by using the U.S. consumer price indices published by the U.S. Bureau of the Census 1993.

We find the most notable differences between the two data sets in return on assets and total firm assets. The Transport data shows lower values in both areas. Also, debt is much more often subordinated in the Transport data. Close to 50% of the bond issues are subordinated, versus under 20% in the Industrial data set. Finally, the debt ratio is much more variable among the Industrial data than the Transport data.

4. Classification methods

Our classification models are OLS, multinomial (linear) discriminant analysis (MDA) with equal priors and with proportional priors, and ordered maximum likelihood (ordered probit and ordered logit). We make all classifications to the category with the highest estimated probability of occurrence. The MDA and ordered probit regression method appear frequently in the bond rating literature. The success of these models varies from one study to another, as shown in Table 1. Ederington (1985) describes most of these techniques.

Table 4

Summary statistics for independent variables

We report summary statistics for the variables used in our forecasting model. The bond rating categories for Transport and for Industrial range from Aaa to B. A rating of B was assigned the value of 0, Ba a value of 1, Baa a value of 2, A a value of 3, Aa a value of 4 and Aaa a value of 5. Interest coverage is defined as net income plus interest expense, divided by interest expense. Debt ratio is defined by total debts divided by total assets. Return on assets is defined as net income divided by total assets. Subordination status is defined as1 if the debt issue has seniority, and 0 otherwise.

Variable	Dataset	Mean	Standard Deviation	Minimum	Maximum
Bond rating	Transport	2.55	1.09	0	5.00
	Industrial	1.86	1.34	0	5.00
Interest coverage	Transport	2.27	8.60	- 2.91	70.8
	Industrial	3.64	7.43	-15.9	56.1
Debt ratio	Transport	0.713	0.120	0.397	0.942
	Industrial	0.800	0.720	0.0892	6.37
Return on assets	Transport	0.00255	0.042	- 0.078	0.102
	Industrial	0.0250	0.096	- 0.308	0.636
Total firm assets[a]	Transport	6670	3390	353	13300
	Industrial	8240	14300	150	121000
Subordination status	Transport	0.483	0.503	0.000	1.00
	Industrial	0.170	0.376	0.000	1.00

[a] Measured in millions of dollars, adjusted for inflation.

In this study we use the ordered logit combining[1] technique of Kamstra and Kennedy (1998) to combine forecasts from ordered probit and MDA (equal priors). Given the same information, different models will have different underlying assumptions and thus will produce different probability estimates. Therefore, we investigate whether combining these different estimates can produce better forecasts than the individual models. The key feature of the methods to be combined (ordered probit and MDA equal priors) is that they are not too similar to each other. Excessive collinearity of forecasts leads to unstable combination weight estimation.[2]

We investigate the performance of the alternative methods using the leaving-one-out method of Lachenbruch and Mickey (1968), also referred to as cross-validation. To implement cross-validation, the model is estimated on all-but-one-observation of the data, and then the estimated model is used to forecast the remaining observation, the hold-out observation. The hold-out forecast is then collected and

[1] This "combining" use of ordered logit is not to be confused with using ordered logit as a classification method in its own right.

[2] Only two forecasts are used for combining in this study because all the methods produce forecasts that are highly collinear. Including three or more forecasts generally leads to problems identifying model parameters, unstable models and unreliable forecasts. Other combinations of two forecasts produce very similar results.

the process repeated, leaving out a different data point each time, until hold-out forecasts for the entire data set are produced.

5. Combining forecasts

In this paper we compare the forecasting performances of several common probability forecasting procedures to the forecasting performances of three methods that combine results from two of these common procedures. The first of these combining methods is simple averaging of probability forecasts, a method that needs no explanation. The second and third are variants of the method of Kamstra and Kennedy (1998), an explanation of which follows.

We begin our explanation of the Kamstra-Kennedy combination method with the binary 0/1 case. For convenience, we refer to zero as a failure, and one as a success. There are two forecast methods, A and B (say probit and MDA), available, each of which produces a probability forecast of a success. We specify the true probability of success for the i^{th} observation to equal a cumulative distribution F evaluated at the index value θ_i. Therefore, the forecast of the probability of success that results from using methods A and B, p_{Ai} and p_{Bi}, can also be expressed as a function of the cumulative distribution F and implicit index values θ_{Ai} and θ_{Bi} so that

$$p_{Ai} = F(\theta_{Ai}) \text{ and } p_{Bi} = F(\theta_{Bi}). \quad (1)$$

The cumulative distribution F is not known. For several pragmatic reasons, Kamstra and Kennedy (1998) use the logit function for the unknown F,

$$prob(success) = e^{\theta}/(1+e^{\theta}). \quad (2)$$

With probability forecasts of success, p_{Ai} and p_{Bi}, we can derive the implicit index values θ_{Ai} and θ_{Bi} from the log odds ratios,

$$\theta_{Ai} = \ln[p_{Ai}/(1-p_{Ai})] \text{ and } \theta_{Bi} = \ln[p_{Bi}/(1-p_{Bi})] \quad (3)$$

The Kamstra-Kennedy method then uses a logit regression with explanatory variables θ_{Ai} and θ_{Bi}. Kamstra and Kennedy (1998) interpret the coefficients on these explanatory variables as the weight placed on the forecasts from methods A and B, respectively. If method A adds no information beyond that provided by method B, but method B provides information beyond that provided by method A, we would expect a weight (regression coefficient from the logit regression) of zero for method A, and one on method B. If the two methods are equally, but differently, informative, we would expect a weight of 0.5 for both. Kamstra and Kennedy do not claim optimality results for this method. They see logit combining as a computationally attractive method for alleviating bias (because it includes an intercept in the logit specification) and for providing data-based weightings.

We now consider the general case for ordered polychotomous data. There is an index θ, which for convenience we call credit-worthiness, which determines classification. We suppose there are J categories, ordered from one, the lowest, to J, the highest. As θ increases and exceeds progressively larger unknown threshold values α_j, j = 1, ...J−1, the classification changes from category j to category j+1. The probability that the i^{th} observation belongs to category j is given by the integral of a logit from α_{j-1}-θ_i to α_j-θ_i. For j = 1 the lower limit is minus infinity and for j = J the upper limit is infinity. α_1 is normalized to be zero. This is the ordered logit model. See Greene (1993) for details. For each observation, each forecasting technique produces the measures $\omega_{ji} = \alpha_j - \theta_i$, j = 1, ..., J−1. We can estimate these measures as

$$\omega_{ji} = \left[\frac{p_{1i} + \ldots\ldots + p_{ji}}{1 - p_{1i} - \ldots\ldots - p_{ji}} \right] \tag{4}$$

where p_{ji} is an estimate of the probability that the i^{th} observation falls in the j^{th} category. Thus, technique A's estimated probability that the i^{th} observation belongs to category j is given by the integral of a standard logit from $\omega_{j-1,i,A}$ to $\omega_{j,i,A}$, where the undefined $\omega_{0,i,A}$ and $\omega_{J,i,A}$ are minus infinity and plus infinity, respectively.

The ordered logit combining method of Kamstra and Kennedy (1998) involves suitably weighting different techniques' $\omega_{j,i}$ values. In this case, combining occurs in the space of integral limits rather than probability or index (θ) space.

For example we suppose there are three categories, ordered from lowest to highest as 1 = junk bonds, 2 = low-grade bonds and 3 = investment-grade bonds. Then for the i^{th} observation,

$$prob(junk) = \frac{e^{\alpha 1-\theta_i}}{1 + e^{\alpha 1-\theta_i}} = \frac{e^{\omega 1i}}{1 + e^{\omega 1i}} \tag{5}$$

$$prob(lowgrade) = \frac{e^{\alpha 1-\theta_i}}{1 + e^{\alpha 1-\theta_i}} \tag{6}$$

$$prob(investmentgrade) = \frac{1}{1 + e^{\alpha 2-\theta_i}} = \frac{1}{1 + e^{\omega 2i}} \tag{7}$$

Probability forecasts from a method like ordered probit or MDA will generate the ω values. The Kamstra and Kennedy (1998) combining method consists of finding, via maximum likelihood, an appropriate weighted average of the competing forecasting techniques' integral limits (ω's).

We suppose there are two forecast methods A and B (say probit and MDA) that produce probability forecasts, and hence ω values. Kamstra and Kennedy (1998) specify that for the i^{th} observation, the contribution to the "combining" logit likelihood function of the two forecast methods is

$$prob(junk) = \frac{e^{\pi_1 + \sum_{l=A,B} \pi_l \omega_{1il}}}{1 + e^{\pi_1 + \sum_{l=A,B} \pi_l \omega_{1il}}} \quad (8)$$

$$prob(lowgrade) = \frac{e^{\pi_2 + \sum_{l=A,B} \pi_l \omega_{2il}}}{1 + e^{\pi_2 + \sum_{l=A,B} \pi_l \omega_{2il}}} - \frac{e^{\pi_1 + \sum_{l=A,B} \pi_l \omega_{1il}}}{1 + e^{\pi_1 + \sum_{l=A,B} \pi_l \omega_{1il}}} \quad (9)$$

$$prob(investmentgrade) = \frac{1}{1 + e^{\pi_2 + \sum_{l=A,B} \pi_l \omega_{2il}}} \quad (10)$$

Kamstra and Kennedy (1998) describe their algorithm as implementable with standard packages, but in fact this is not so. Standard packages do not allow different values for the variables in the logit across categories (the ω_{jil}, j = 1,2 must not vary over j). However, in this example they do vary. We refer to this below as the KK method. Although we could code this as a MLE problem, we want to consider a modification of Kamstra and Kennedy algorithm that can be implemented with standard logit packages. This modification specifies that for the i^{th} observation

$$prob(junk) = \frac{e^{\pi_1 + \sum_{l=A,B} \sum_{j=1}^{2} \pi_{j,l} \omega_{jil}}}{1 + e^{\pi_1 + \sum_{l=A,B} \sum_{j=1}^{2} \pi_{j,l} \omega_{jil}}} \quad (11)$$

$$prob(junk) = \frac{e^{\pi_2 + \sum_{l=A,B} \sum_{j=1}^{2} \pi_{j,l} \omega_{jil}}}{1 + e^{\pi_2 + \sum_{l=A,B} \sum_{j=1}^{2} \pi_{j,l} \omega_{jil}}} - \frac{e^{\pi_1 + \sum_{l=A,B} \sum_{j=1}^{2} \pi_{j,l} \omega_{jil}}}{1 + e^{\pi_1 + \sum_{l=A,B} \sum_{j=1}^{2} \pi_{j,l} \omega_{jil}}} \quad (12)$$

$$prob(investmentgrade) = \frac{1}{1 + e^{\pi_2 + \sum_{l=A,B} \sum_{j=1}^{2} \pi_{j,l} \omega_{jil}}} \quad (13)$$

That is, we use the ω's for each category and method as logit regressors. The ω's are no different from ordinary explanatory variables that are positively correlated with the category level, just as profit levels and total firm assets are with bond classification level. This model can be estimated as a standard ordered logit with the π_1 and π_2 parameters playing the role of the unknown threshold values.

The disadvantage of this method is that it consumes more degrees of freedom because there are many more parameters to estimate. In addition to being easy to use, the advantage of this method is that the ω_{jil} values for one category are informative for all categories, and this information is available in this formulation. We refer to this below as the m-KK method, for the modified Kamstra-Kennedy method.

The ordered logit combining method uses in-sample probability forecasts from MDA and ordered probit, transformed into ω values, to form a rule for combining the hold-out sample forecasts. We then repeat this procedure in a cross-validated manner to produce combined hold-out sample forecasts analogous to the hold-out sample forecasts from MDA and ordered probit. All forecasts will be compared only on the hold-out sample performance, because out-of-sample performance can best determine the forecasting ability of the various methods on new data sets.

6. Performance evaluation rules

We use two scoring rules to measure the performance of the individual classification models and different combining methods, the error rate, and the probability score.

The error rate estimates the percentage of observations that a model predicts incorrectly. We calculate it by taking the ratio of the number of incorrect predictions and the number of observations in the data set. The better the predictive power of a model, the lower is the ratio.

The probability score of Brier (1950) calculates the mean squared errors of the probabilistic forecasts. If there are J categories and N observations, event $e_{i,j}$ takes on the value 1 if the i^{th} observation belongs to category j, zero otherwise. The forecasting method predicts that the probability of landing in category j for the i^{th} observation is $p_{i,j}$. The probability score has the form:

$$\sum_{i=1}^{N} \sum_{j=1}^{J} (e_{i,j} - p_{i,j})^2/N. \tag{14}$$

To illustrate the usefulness of this score, we use an example for the binary case, where the dependent variable takes on the value of one or zero. We suppose method A always places the probabilities of 0.49 on one category and 0.51 on the other. Method B always places the probability of 0.99 on the correct category whenever method A is correct. When the two methods are incorrect, method B places the same probabilities as method A, 0.49 on the correct category. Methods A and B have the same error rate. However, method B has a much better (lower) probability score, reflecting its higher degree of confidence in its correct forecasts. The best possible score for a method (zero) occurs when it forecasts with certainty that the actual event will occur. The worst possible score for a method (two) occurs when it forecasts that the actual event will occur with zero probability.

7. Estimation results

In-sample estimation results for several of our models appear in the Appendix. See the Appendix for a short discussion of these results. Since we are interested in forecasting, we focus on the out-of-sample estimation results found in Table 5.

Table 5 shows the error rate and the probability score of each method for the two data sets, Transport and Industrial. We use the ordered probit and MDA (equal priors) methods as inputs to the combination forecast methods.

We draw two general conclusions from these results, qualified only by some isolated exceptions, such as that the MDA (proportional priors) method has a low error rate score in the Transport data. First, the combining methods outperform the

Table 5

Error rates and probability scores

We report summary statistics on the forecast errors for our various forecasting methods and data sets. The error rate estimates the percentage of observations that a model predicts incorrectly. We calculate it by taking the ratio of the number of incorrect predictions and the number of observations in the data set. The better the predictive power of a model, the lower is the ratio. The probability score calculates the mean squared errors of the probabilistic forecasts. The better the predictive power of a model, the lower is the probability score. If there are J categories and N observations, event $e_{i,j}$ takes on the value 1 if the i^{th} observation belongs to category j, zero otherwise. The forecasting method predicts that the probability of landing in category j for the i^{th} observation is $p_{i,j}$. The probability score has the form:

$$\sum_{i=1}^{N} \sum_{j=1}^{J} (e_{i,j} - p_{i,j})^2 / N.$$

Method	Transport Data		Industrial Data	
	Error Rate	Probability Score	Error Rate	Probability Score
Ordered Probit	0.461	0.592	0.543	0.670
Ordered Logit	0.416	0.596	0.525	0.655
MDA (Equal Priors)	0.551	0.573	0.543	0.712
MDA (Proportional Priors)	0.371	0.531	0.581	0.708
OLS	0.472	0.944	0.611	1.223
Simple Average Combination	0.404	0.524	0.543	0.670
Ordered Logit Combination (KK)	0.382	0.515	0.547	0.671
Ordered Logit Combination (m-KK)	0.371	0.533	0.472	0.637

non-combining methods, consistent with the forecast combining literature. Second, the m-KK method is the best of the combining methods.

Since many of the measures in Table 5 are similar, we wish to see if the differences between the combining methods and their input forecasts are significant. We use paired sample t-tests to make this comparison We let E_{Ai} and E_{Bi} be the error series for forecasting methods A and B. $E_{Ai} = 0$ if forecast method A correctly predicts the bond rating for the i^{th} observation, one otherwise. Our test is a standard paired-sample t-test for mean greater than 0 of E_{Ai} -E_{Bi}. This t-test is uniformly most powerful among unbiased tests only under limited conditions, such as normality of E_{Ai} -E_{Bi}, a condition which will not be satisfied in the setting here as the E's are 0/1 variables. However, this test has an asymptotic normal distribution under weak conditions, permitting non-normality and some dependence and heterogeneity. To show this result we appeal to the appropriate central limit theorem. See Lehmann (1975) and Gastwirth and Rubin (1971).

Table 6 shows the *p*-values (rounded to three decimal places) associated with the paired *t*-tests for significant difference in means of the error rates. Table 7 shows the *p*-values associated with the paired t-tests for significant difference in means of the probability scores. To keep results manageable we report results only for the

Table 6

Paired t-tests of significant difference in error rates

All entries are probability-values of the t-tests. A p-value less than α means that the method along the row has a mean error rate that is significantly lower at the α level than the method along the column. The error rate estimates the percentage of observations that a model predicts incorrectly. We calculate it by taking the ratio of the number of incorrect predictions and the number of observations in the data set.

	Transport Data			Industrial Data		
Method	MDA Equal Priors	Ordered Probit	Simple Average	MDA Equal Priors	Ordered Probit	Simple Average
Ordered Probit	0.100	-	-	0.500	-	-
Simple Average	< 0.0005	0.160	-	0.500	0.500	-
m-KK Logit Combination	< 0.0005	0.040	0.230	0.010	< 0.0005	< 0.0005

Table 7

Paired t-tests of significant difference in probability scores

All entries are probability-values of the t-tests. A p-value less than α means that the method along the row has a probability score that is significantly lower at the α level than the method along the column. The probability score calculates the mean squared errors of the probabilistic forecasts. The better the predictive power of a model, the lower is the probability score. If there are J categories and N observations, event $e_{i,j}$ takes on the value 1 if the i^{th} observation belongs to category j, zero otherwise. The forecasting method predicts that the probability of landing in category j for the i^{th} observation is $p_{i,j}$. The probability score has the form:

$$\sum_{i=1}^{N} \sum_{j=1}^{J} (e_{i,j} - p_{i,j})^2 / N.$$

	Transport Data			Industrial Data		
Method	MDA Equal Priors	Ordered Probit	Simple Average	MDA Equal Priors	Ordered Probit	Simple Average
Ordered Probit	0.620	-	-	0.020	-	-
Simple Average	0.070	0.020	-	< 0.0005	0.370	-
m-KK Logit Combination	0.260	0.140	0.580	< 0.0005	< 0.0005	< 0.0005

m-KK method, the Simple Average method, and their input forecast methods. Adding other methods to Table 6 and Table 7 would not change our overall conclusions.

A p-value in Table 6 or Table 7 less than 0.05 indicates that the method along the row has a mean error rate that is significantly lower at the 5% level than the method along the column. For example, Table 5 shows that ordered probit has a mean error rate that is lower than MDA (equal priors) for both data sets, and considerably lower for the Transport data set. Table 6 shows that this difference is significant at the 10% level (exactly) for the Transport data set, but not for the Industrial data set.

To calculate the p-value for the hypothesis that the error rate associated with the method along a column is significantly lower than the method along a row, we subtract the tabled p-value for that row and column from one. For example, in Table 6, the p-value that the Simple Average method error rate for the Transport data set is significantly lower than the error rate for the ordered probit is (1-0.16) =0.84.

The results reported in Table 6 and Table 7 support the conclusions drawn earlier using the results reported in Table 5. First, the combining methods outperform their input forecasts, in several instances by a margin that is significant at the 5% level. Second, the m-KK method significantly outperforms its combining competition (the Simple Average method) in several instances, but the reverse is never the case. The m-KK method does particularly well using the Industrial data, the larger data set. This is not surprising. The m-KK method uses the data to estimate its weighting parameters, but the Simple Average method involves no parameter estimation. With an insufficient number of observations, any scheme that relies on parameter estimation is bound to suffer relative to the Simple Average method.

In summary, we see that the m-KK logit combining method performs well when there is a large data set available, and that with a small data set, the m-KK logit combining method still does not show any significant (or substantive) disadvantage compared to the Simple Average method. This makes a strong case for combining qualitative forecasts of bond ratings, as well as a strong case for combining using a regression-based m-KK logit method rather than a Simple Average method.

8. Conclusion

In this paper we use ordered logit regression combining methods to form a consensus forecast from differing individual forecasts to predict bond ratings in the transportation and industrial sectors. We investigate two alternative logit combining methods, one proposed in Kamstra and Kennedy (1998), and a modification of this Kamstra-Kennedy method. We found that the modified Kamstra-Kennedy method, which is implementable with standard logistic estimation software, is preferable over the Kamstra and Kennedy (1998) method.

Our combining (consensus) results are consistent with those of other researchers, that in general combined forecasts outperform their input forecasts. Furthermore, we show that in a data set of 265 observations, an ordered-logit forecast combination of bond ratings yields statistically significant, quantitatively meaningful improvements in forecasting over the traditional combining method of averaging input probability forecasts. We also show that in a data set of only 89 observations this ordered-logit combining method continues to perform well, albeit with a fall in its relative performance. We conclude that, with an adequate number of observations, forecasting in the context of ordered qualitative outcomes is best undertaken using a variant of the ordered logit combining technique introduced by Kamstra and Kennedy (1998).

Appendix A

Table 8 shows that the in-sample results from OLS, ordered probit, and ordered logit are consistent with one another. The results show that all variables except the interest coverage variable are significant. Of the significant variables, all have sensible signs, negative for subordination status and debt ratio, and positive for total firm assets and return on assets. Bond ratings are generated by multiplying these coefficients by the value taken on for subordination status, total firm assets, interest coverage, debt ratio, and return on assets, and summing and calculating the category. For the Transport data case and the OLS model, we multiply the coefficients by the mean values of the corresponding explanatory value, 6670 for total firm assets, 2.27 for interest coverage, 0.713 for debt ratio, and 0.00255 for return on assets, and set subordination status to one. This method yields a bond rating forecast of 2.24, which would round to a category 2, or a Baa, rating. The mean rating rating is 2.55, between a Baa and A rating. See Greene (1993) for implementation details for the maximum likelihood methods.

Table 9 presents the modified Kamstra-Kennedy results (the coefficient estimates, π's). These coefficients, many of which are strongly significant, are the logit estimates on the ω's. For a description, see Section 5 on combining forecasts. The coefficients on the ω's from the ordered probit (π_{ORD}) are much less significant than those found on the ω's from the MDA (equal priors) procedure (π_{MDA}). This result appears to be an artifact of the high correlation (unreported) between ω's from the ordered probit. The correlation between ω's from the MDA procedure is between 0.65 and 0.9, but for ordered probit these numbers range from over 0.9 to 0.98. Unfortunately, there is no natural interpretation for the π's of the m-KK method, unlike the KK method. The ω's themselves can range from negative to positive infinity, theoretically.

Table 10 reports the values of the ω's corresponding to the estimated coefficients (the π's reported in Table 9). Section 5 describes how the combining model can be used for forecasting.

Table 8

Estimated coefficients from OLS, ordered probit and ordered logit

Ordered Probit and Ordered Logit: p_{ji} (bond rating for observation i = category j) = $F(\alpha_j - X_i\beta)$-$F(\alpha_{j-1} - X_i\beta)$, where p_{ji} is an estimate of the probability that the ith observation falls in the jth category, the Xs include constant terms (threshold values). For Ordered Probit the function F is the cumulative normal. For Ordered Logit the function F is the logit, F(x) = 1/(1+exp(-x)). Interest coverage is defined as net income plus interest expense, divided by interest expense. Debt ratio is defined by total debts divided by total assets. Return on assets is defined as net income divided by total assets. Subordination status is defined as1 if the debt issue has seniority, and 0 otherwise.

	Parameter Estimate (Standard Deviation)					
	Transport Data			Industrial Data		
Variable	OLS	Ordered Probit	Ordered Logit	OLS	Ordered Probit	Ordered Logit
Intercept	5.52***			1.99***		
	(.657)			(0.114)		
Intercept B[a]					−1.26***	−2.67***
					(0.153)	(0.324)
Intercept Ba	−6.52***	−11.4***	0.620***	−1.52***		
	(1.30)	(2.33)	(0.0946)	(0.277)		
Intercept Baa	2.63***	−6.80***	1.69***	0.363		
	(0.364)	(2.21)	(0.133)	(0.258)		
Intercept A	3.68***	−4.99***	2.87***	2.50***		
	(0.424)	(1.99)	(0.173)	(0.337)		
Intercept Aa	5.03***	−2.60***	3.92***	4.56***		
	(0.558)	(1.97)	(0.261)	(0.549)		
Subordination Status	−0.619***	−1.20***	−2.05***	−1.50***	−1.81***	−3.11***
	(0.154)	(0.293)	(0.524)	(0.174)	(0.223)	(0.413)
Total firm assets	5.06E-5***	1.19E-4***	2.00E-4***	2.94E-5***	3.11E-5***	6.00E-5***
	(2.44E-5)	(4.50E-5)	(7.80E-5)	(1.50E-6)	(4.89E-6)	(9.77E-6)

(continued)

Table 8 *(continued)*

Estimated coefficients from OLS, ordered probit and ordered logit

	Parameter Estimate (Standard Deviation)					
	Transport Data			Industrial Data		
Variable	OLS	Ordered Probit	Ordered Logit	OLS	Ordered Probit	Ordered Logit
Interest coverage	–0.0030	–0.015	–0.032	0.011	0.012	0.0017
	(0.010)	(0.016)	(0.029)	(0.010)	(0.011)	(0.020)
Debt ratio	–4.22***	–6.35***	–11.0***	–0.364***	–0.460***	–1.48***
	(0.895)	(1.66)	(2.92)	(0.098)	(0.112)	(0.276)
Return on assets	7.89***	17.7***	33.1***	5.30***	6.56***	17.0***
	(2.67)	(4.90)	(8.87)	(0.865)	(1.07)	(2.39)

*** Indicates statistical significance at the 0.01 level.
** Indicates statistical significance at the 0.05 level.
* Indicates statistical significance at the 0.10 level.
[a]: For the Transport data the single observation with a B rating is merged with the Ba class.

Table 9

Estimated coefficients from m-KK ordered logit combining

$\pi_{MDA,j}$ is the coefficient on the probability index ω_j from MDA (equal priors) for the j^{th} bond category, and $\pi_{ORD,j}$ is the coefficient on the probability index ω_j from ordered probit for the j^{th} bond category. Note that the category Aaa is the terminal category and thus does not have coefficients estimated for it.

	Parameter Estimate (Standard Deviation)	
Variable	Transport Data[a]	Industrial Data
Intercept B		-11.4 (19.8)
Intercept Ba	25.2 (25.4)	-10.2 (19.8)
Intercept Baa	32.2 (25.8)	-7.97 (19.8)
Intercept A	34.4*(25.8)	-5.42 (18.8)
Intercept Aa	38.3* (26.0)	-3.1 (19.8)
$\pi_{MDA,B}$		-2.55*** (0.99)
$\pi_{MDA,Ba}$	1.17*** (0.369)	-1.09 (0.98)
$\pi_{MDA,Baa}$	2.40*** (0.566)	1.49*** (0.55)
$\pi_{MDA,A}$	-1.43*** (0.357)	0.88*** (0.32)
$\pi_{MDA,Aa}$	1.97*** (0.465)	-0.086 (0.183)
$\pi_{ORD,B}$		-7.02 (44.4)
$\pi_{ORD,Ba}$	3.46 (4.02)	32.5 (84.5)
$\pi_{ORD,Baa}$	-27.8 (23.6)	-44.3 (55.1)
$\pi_{ORD,A}$	41.1 (33.4)	21.1* (13.7)
$\pi_{ORD,Aa}$	-17.6* (13.1)	-2.22 (3.40)

*** Indicates statistical significance at the 0.01 level.
** Indicates statistical significance at the 0.05 level.
* Indicates statistical significance at the 0.10 level.
[a]: For the Transport data the single observation with a B rating is merged with the Ba class.

Table 10

Summary statistics on the regressors (ω terms) of the m-KK method

We report summary statistics on the regressors used in the m-KK method, which are transformations of the probability forecasts from the MDA and ordered probit models. We calculate the I[th] observation of $\omega_{MDA,j}$ (i.e., $\omega_{j,I}$) as

$$\omega_{ji} = \left[\frac{p_{1i} + \ldots\ldots + p_{ji}}{1 - p_{1i} - \ldots\ldots - p_{ji}}\right]$$

where p_{ji} is the MDA's probability forecast for the I[th] observation to be in the j[th] category.

Variable	Mean	Standard Deviation	Minimum	Maximum
Panel A: Transport Data				
$\omega_{MDA,B}$	−10.8	6.56	−27.5	6.57
$\omega_{MDA,Ba}$	−10.2	6.12	−26.2	−0.801
$\omega_{MDA,Baa}$	−3.37	4.03	−15.8	7.46
$\omega_{MDA,A}$	−5.15	4.29	−16.9	1.67
$\omega_{MDA,Aa}$	−0.270	3.45	−13.1	9.07
$\omega_{ORD,B}$	0.802	2.87	−5.67	8.90
$\omega_{ORD,Ba}$	3.13	4.68	−11.8	18.16
$\omega_{ORD,Baa}$	3.13	3.36	−2.94	13.5
$\omega_{ORD,A}$	6.09	4.61	−3.42	21.0
$\omega_{ORD,Aa}$	6.97	4.69	−0.44	21.1
Panel B: Industrial Data				
$\omega_{MDA,B}$	−2.46	2.16	−13.1	3.00
$\omega_{MDA,Ba}$	−1.72	1.80	−9.93	4.93
$\omega_{MDA,Baa}$	−0.747	2.31	−11.3	6.93
$\omega_{MDA,A}$	−0.633	1.75	−7.70	6.69
$\omega_{MDA,Aa}$	0.485	2.22	−11.0	8.24
$\omega_{ORD,B}$	1.12	1.96	−4.64	10.4
$\omega_{ORD,Ba}$	1.67	2.25	−10.2	10.3
$\omega_{ORD,Baa}$	3.33	2.56	−2.16	15.5
$\omega_{ORD,A}$	4.81	3.10	−8.74	17.2
$\omega_{ORD,Aa}$	5.97	3.33	−0.422	21.2

References

Bates, J. and C. W. Granger, 1969. The combination of forecasts, *Operational Research Quarterly* 20, 451-468.

Belkaoui, A., 1980. Industrial bond ratings: A new look, *Financial Management* 9, 44-51.

Bopp, A. E., 1985. On combining forecasts: Some extensions and results, *Management Science* 31, 1492-1498.

Bordley, R., 1986. Technical note: Linear combination of forecasts with an intercept: A Bayesian approach, *Journal of Forecasting* 5, 243-429.

Brier, G., 1950. Verification of forecasts expressed in terms of probability, *Monthly Weather Review* 78, 1-3.

Bunn, D., 1975. A Bayesian approach to linear combination of forecasts, *Operational Research Quarterly* 26, 325-329.

Bunn, D., 1981. Two methodologies for the linear combination of forecasts, *Journal of the Operational Research Society* 32, 213-222.

Clemen, R. T., 1986. Linear constraints and the efficiency of combined forecasts, *Journal of Forecasting* 5, 31-38.

Clemen, R. T., 1989. Combining forecasts: A review and annotated bibliography, *International Journal of Forecasting* 5, 559-583.

Clemen, R. T. and R. L. Winkler, 1986. Combining economic forecasts, *Journal of Business and Economic Statistics* 4, 39-46.

Datta, S., M. Iskandar-Datta, and A. Patel, 1997. The pricing of initial public offers of corporate straight debt, *Journal of Finance* 52, 379-96.

Donaldson, R. G. and M. Kamstra, 1996. Forecast combining with neural networks, *Journal of Forecasting* 15, 49-61.

Donaldson, R. G. and M. Kamstra, 1999. Neural network forecast combining with interaction effects, *Journal of Franklin Institute* 336, 227-236.

Ederington, L. H., 1985. Classification models and bond ratings, *The Financial Review* 20, 237-261.

Feather, P. M. and M. S. Kaylen, 1989. Conditional qualitative forecasting, *American Journal of Agricultural Economics* 71, 195-201.

Fan, D. K., K.-N. Lau, and P.-L. Leung, 1996. Combining ordinal forecasts with an application in a financial market, *Journal of Forecasting* 15, 37-48.

Followill, R. A. and T. Martell, 1997. Bond review and rating change announcements: An examination of informational value and market efficiency, *Journal of Economics and Finance* 21, 75-82.

Fox, J., 1984. *Linear Statistical Models and Related Methods* (Wiley, New York, NY).

Gastwirth, J. L. and H. Rubin, 1971. The behavior of robust estimators on dependent data, Pudue University Statistics Mimeo 197.

Gentry, J. A., D. T. Whitford, and P. Newbold, 1988. Predicting industrial bond ratings with a probit model and funds flow components, *The Financial Review* 23, 269-286.

Goh, J. C. and L. H. Ederington, 1993. Is a bond rating downgrade bad news, good news, or no news for stockholders? *Journal of Finance* 48, 2001-2008.

Granger, C. W., 1989. Invited review: Combining forecasts-twenty years later, *Journal of Forecasting* 8, 167-173.

Granger, C. W. and R. Ramanathan, 1984. Improved methods of combining forecasts, *Journal of Forecasting* 3, 197-204.

Greene, W. H., 1993. *Econometric Analysis* (Prentice Hall, Englewood Cliffs, NJ).

Gunter, S. I. and C. Aksu, 1989. N-Step combinations of forecasts, *Journal of Forecasting* 8, 253-267.

Hand, J. R. M., R. W. Holthausen, and R. W. Leftwich, 1992. The effect of bond rating agency announcements on bond and stock prices, *Journal of Finance* 47, 733-752.

Harrald, P. and M. Kamstra, 1997. Evolving artificial neural networks to combine financial forecasts, *IEEE Transactions on Evolutionary Computation* 1, 40-52.

Hallman, J. and M. Kamstra, 1989. Combining algorithms based on robust estimation techniques and co-integrating restrictions, *Journal of Forecasting* 8, 189-198.

Impson, C. M., I. Karafiath, and J. L. Glasscock, 1992. Testing beta stationarity across bond rating changes, *The Financial Review* 27, 607-618.

Kamstra, M. and P. Kennedy, 1998. Combining qualitative forecasts using logit, *International Journal of Forecasting* 14, 83-93.

Kaplan, R. S. and G. Urwitz, 1979. Statistical models of bond ratings: A methodological inquiry, *Journal of Business* 52, 231-261.

Kao, C. and C. Wu, 1990. Two-step estimation of linear models with ordinal unobserved variables: The case of corporate bonds, *Journal of Business and Economic Statistics* 8, 317-325.

Lachenbruch, P. A. and M. R. Mickey, 1968. Estimation of error rates in discriminant analysis, *Technometrics* 10, 1-11.

Lehmann, E. L., 1975. *Non-Parametrics* (Holden-Day, San Francisco, CA).

Mahmoud, E., 1984. Accuracy in forecasting: A survey, *Journal of Forecasting* 3, 139-159.

Makridakis, S. and R. L. Winkler, 1983. Averages of forecasts: Some empirical results, *Management Science* 29, 987-996.

Moody's Investors Service, Inc., 1988-1993. *Moody's Transportation Manual* (Moody's Investors Service, Inc., New York, NY).

Moody's Investors Service, Inc., 1993. *Moody's Industrial Manual* (Moody's Investors Service, Inc., New York, NY).

Moon, C. G. and J. G. Stotsky, 1993a. Municipal bond rating analysis, *Regional Science and Urban Economics* 23, 29-50.

Moon, C. G. and J. G. Stotsky, 1993b. Testing the difference between the determinants of Moody's and Standard's and Poor's ratings, *Journal of Applied Econometrics* 8, 51-69.

Newbold, P. and C. W. Granger, 1974. Experience with forecasting univariate time series and the combination of forecasts, *Journal of Royal Statistical Society: Series A* 137, 131-149.

O'Neal, E. S., 1998. Why electric utility stocks are sensitive to interest rates, *The Financial Review* 33, 147-61.

Pinches, G. E. and K. A. Mingo, 1973. A multivariate analysis of industrial bond ratings, *Journal of Finance* 28, 1-17.

Pogue, T. F. and R. M. Soldofsky, 1969. What's in a bond rating, *Journal of Financial and Quantitative Analysis* 4, 201-228.

Reiter, S. A. and D. A. Ziebart, 1991. Bond yields, ratings and financial information: Evidence from public utility issues, *The Financial Review* 26, 45-73.

Sanders, F., 1963. Subjective probability forecasting, *Journal of Applied Meteorology* 2, 191-201.

Staël von Holstein, C.-A. S., 1971. An experiment in probabilistic weather forecasting, *Journal of Applied Meteorology* 10, 635-645.

Staël von Holstein, C.-A. S., 1972. Probabilistic forecasting: An experiment related to the stock market, *Organizational Behavior and Human Performance* 8, 139-158.

Terza, J. V., 1985. Ordinal probit: A generalization, *Communications in Statistics, Theory and Methods* 14, 1-12.

Trenkler, G. and E. P. Linski, 1986. Note: Linear constraints and the efficiency of combined forecasts, *Journal of Forecasting* 5, 197-202.

U.S. Bureau of the Census, 1993. *Statistical Abstract of the United States* 113th ed. (U.S. Bureau of the Census, Washington, DC).

Winkler, R. L. and S. Makridakis, 1983. The combination of forecasts, *Journal of Royal Statistical Society: Series A* 146, 150-157.

Winkler, R. L., A. Murphy, and R. Katz, 1977. The consensus of selective probability forecasts: Are two, three, ... heads better than one?, In *Preprint Volume Fifth Conference on Probability and Statistics*, 57-62.