Review of Pacific Basin Financial Markets and Policies
Vol. 7, No. 2 (2004) 153–172

Market-Based Evaluation for Models to Predict Bond Ratings

Konan Chan
Department of Finance, College of Management
National Taiwan University, Taipei 10660, Taiwan
kchan@mba.ntu.edu.tw

Narasimhan Jegadeesh
Goizueta Business School, Emory University
1300 Clifton Road, Atlanta, GA 30322-2722, USA
Narasimhan_Jegadeesh@bus.emory.edu

Previous studies have examined different statistical models to predict corporate bond ratings. However, these papers use agency ratings as the benchmark to assess models and ignore the evidence that agency ratings may not be accurate in a timely manner. In this paper, we propose a new approach which incorporates ex-post bond returns to evaluate rating prediction models. Relative rating strength portfolios, formed by buying under-rated bonds with agency ratings lower than model ratings and selling over-rated bonds with agency ratings higher than model ratings, are employed to test the performance of different statistical models in rating predictions. Our results show that one version of multiple discriminant analysis model can generate a statistically significant abnormal return of 5% over a 5-year horizon. The ordered probit model which is believed to possess theoretical advantages in classifying bonds does not perform better. This suggests that using traditional measures to evaluate models can be misleading. The existence of a profitable trading strategy also raises the concern of market efficiency in the corporate bond market.

Keywords: Bond rating prediction; relative rating strength portfolio; bond trading strategy; bond market efficiency.

1. Introduction

There is a bulk of literature examining the performance of various statistical models in corporate bond rating predictions.[1] These previous studies

[1] Horrigan (1966) starts this line of research by using an ordinary least squares regression model. Pinches and Mingo (1973, 1975) apply the multiple discriminant analysis model to improve the statistical fit. Ederington (1985) further investigates the performance of different models in bond rating predictions.

assume that agency ratings provided by Moody's or S&P are good proxies of default risks and therefore use them as the benchmark to evaluate the rating prediction models. However, the evidence that agency ratings may not be accurate in a timely fashion is generally overlooked. For example, Weistein (1977) finds that bond price changes are fully anticipated during the period of 18 to seven months preceding the bond rating revisions made by Moody's, a result further supported by Pinches and Singleton (1978) who examine the stock returns. Moreover, Hite and Warga (1997) show that the cumulative abnormal bond return within six months prior to rating changes is about two to ten times larger than that in the announcement month.[2] In other words, the bond market has anticipated most of the rating process long before the rating change announcements. This pre-revision information leakage suggests that agency ratings may not be unbiased estimators of credit risks at each point of time. Therefore, the assumption in previous rating prediction research is not valid.

One possible reason why agency ratings may be biased is that Moody's and S&P cannot afford a day-to-day monitoring on thousands of corporate bonds in the market. For example, Ederington and Yawitz (1987) survey the bond rating process and find that there are insufficient analysts in Moody's and S&P to rate corporate bonds. Although both rating agencies claim to have the continuous information gathering from issuers, they tend to meet issuers only once a year. As described in Howe (1995) and Standard & Poor's (1998), the agency rating process is generally conservative and time-consuming. As a result, the rating revisions by rating agencies are usually delayed, leading to information leakage prior to rating change announcements.[3]

In this paper, we propose a new approach to evaluate models in predicting bond ratings. We assume that the market timely incorporates all relevant information through the ratings prediction models to form an independent assessment about default risks of bonds. To the extent that rating agencies tend to delay the rating revisions, if the prediction model can capture

[2]Holthausen and Leftwich (1986) find a similar result in the stock market. Stocks returns in the period of 15 to four months prior to rating changes are much more significant and larger than those in the three months after the rating revisions.

[3]An alternative explanation for biased agency ratings is that there may exist rating biases associated with firm characteristics which cause some bonds to consistently earn higher or lower returns than other issues. The parallel argument can be found in the stock market. For example, Fama and French (1992) find that stocks with high book-to-market ratios and small market-capitalization tend to have higher returns.

the true rating process, there will be subsequent rating changes for any mis-rated bonds whose model ratings are unequal to agency ratings. We therefore form a relative rating strength portfolio at the time of rating predictions by buying under-rated bonds whose agency ratings are worse than model ratings, and selling over-rated bonds whose agency ratings are better than predicted ratings. Since the previous literature shows that there exist abnormal returns prior to rating changes and these return drifts may last over a long horizon,[4] the relative rating strength portfolio should be able to earn abnormal profits in the long-run. Based on the long-run performance of relative rating strength portfolios, we can evaluate the power of different models in classifying corporate bonds. If a statistical model can identify relative rating strength portfolios which generate higher returns than portfolios based on others, this model is more powerful in predicting true bond ratings. Our evaluation criterion is superior to the traditional measure used in earlier studies because we incorporate ex-post bond returns into the evaluation and do not assume that agency ratings are always accurate.

Using 4,474 industrial bonds with Moody's ratings of 415 firms, we perform rating predictions by four statistical models — multiple discriminant analysis (MDA), multiple discriminant analysis with the cross-validation holdout procedure (MDA-C), ordered probit (Probit), and ordered probit with the stepwise variable selection (Probit-S). Our empirical results show that 86% of agency ratings can be correctly replicated by MDA, 75% by MDA-C, 79% by Probit, and 75% by Probit-S based on the traditional accuracy rate measure. As to our proposed approach, only the portfolios from MDA-C can generate significant abnormal returns in the long-run. This supports the notion that the ordered probit model which is documented to possess the theoretical advantages in bond rating predictions does not outperform the multiple discriminant analysis model empirically. Since MDA-C outperforms MDA in our new approach but not in the traditional one, the result also suggests that the evaluation criterion used in previous studies can be misleading. Moreover, to the extent that we find a profitable trading strategy through MDA-C, the corporate bond market is not efficient in processing publicly available information.

This paper contributes to the existing literature in four ways. First, we provide an alternative and better approach by incorporating the ex-post market information to evaluate rating prediction models and document the

[4]Dichev and Piotroski (2001) find significant abnormal stock returns over three years after rating revisions.

potential biases that earlier papers have ignored. Second, we not only examine the performance of different models in bond classifications, but also try to identify some profitable bond trading strategies. Since the ratings prediction literature does not provide any investment implications, this paper attempts to fill this gap by making a connection between rating predictions and corporate bond trading strategies. Third, our results relate to the corporate bond market efficiency, which has received much less attention in the literature. Finally, this paper facilitates our understanding of the relative performance among different rating prediction models since very few previous papers apply more than one model on the same data.

The rest of the paper is organized as follows. Section 2 reviews the models used in previous research to predict bond ratings. Section 3 describes the data and the methodology is presented in Sec. 4. Section 5 shows empirical results and Sec. 6 concludes.

2. Literature Review

Previous studies employ different statistical models to predict bond ratings. Three major models have been examined in the literature, namely, ordinary least squares, multiple discriminant analysis, and ordered probit models. This section reviews these papers and introduces the benefits and restrictions associated with each model.

2.1. *Ordinary least squares model*

Horrigan (1966) pioneers the bond classification research by using ordinary least square (OLS) regressions. He assigns consecutive integers to represent different bond ratings. With the usage of only five financial ratios and one dummy variable for the subordination status, he finds that more than half of the bond ratings can be replicated by his simple OLS model.

However, there are two drawbacks associated with the OLS model in bond rating predictions. First, as pointed out in Kaplan and Urwitz (1979), the consecutive integer dependent variable in OLS assumes that the risk differential between AA and A bonds is the same as that between BB and B bonds. This assumption is clearly not valid in the nature of bond ratings. Second, Mckelvey and Zavoina (1975) illustrate that the error term of an OLS regression does not have a zero mean or constant variance when the dependent variable is ordinal. This indicates that the normality assumption of OLS regressions is violated when assigning ratings as consecutive integers.

2.2. *Multiple discriminant analysis model*

An alternative approach to predict bond ratings is the multiple discriminant analysis (MDA) model. MDA is a statistical technique to classify observations into different groups by maximizing the ratio of between-group variance to within-group variance. However, MDA is not theoretically appropriate in bond rating classifications since it does not consider the ordinal nature of bond ratings. Moreover, MDA requires a strong assumption that independent variables in the model have to follow a multivariate normal distribution. Both Eisenbeis (1977) and Pinches (1978) argue that financial ratios generally do not follow the univariate normal distribution, not even mention multivariate normal. Therefore, though an MDA model avoids the disadvantages associated with an OLS model, it brings new problems to bond rating predictions.[5]

Nevertheless, the MDA model is the most popular rating prediction model in the literature. Pinches and Mingo (1973) first apply the MDA model to bond rating classifications, and demonstrate that about 60% of bonds in their sample can be correctly classified. A bulk of subsequent studies follows the same way to improve the model. The prediction accuracy rate of these papers is about 70%, suggesting that the theoretical disadvantages of MDA do not show up empirically.

2.3. *Ordered probit model*

Introduced by Mckelvey and Zavoina (1975), the ordered probit (Probit) model is designed to solve problems with the ordinal nature. Since it assumes that dependent variables are ordinal and does not require independent variables to be multivariate normal, Probit takes care of the special feature of bond ratings and has theoretical advantages over MDA in predicting bond ratings. Moreover, Probit is flexible in choosing the interval between two groups to best fit the data. Therefore, Probit is also superior to OLS by avoiding the problem of fixed interval between two adjoining rating classes.

[5]The other problem of using the MDA model is the choice between linear and quadratic models. Linear MDA requires variance-covariance matrices (or dispersion matrices) within different groups to be equal while quadratic MDA does not. Because Pinches and Mingo (1975) and Pinches (1978) provide the evidence that dispersion matrices are unequal, quadratic MDA should be the best choice for bond rating predictions. However, Lachenbruch *et al.* (1974) find that when the non-normality is present, quadratic MDA performs worse than linear MDA even dispersion matrices are unequal.

In spite of its suitable application to bond classifications, Probit does not consistently dominate OLS and MDA in the previous literature. For example, Kaplan and Urwitz (1979) compare Probit with OLS and find that OLS slightly outperforms Probit. Wingler and Watts (1982) document that the accuracy rate of Probit is lower than that of MDA in predicting bond rating changes. These results suggest that the appealing theoretical features of Probit do not guarantee a better prediction power. On the contrary, Ederington (1985) shows the outperformance of Probit over OLS and MDA. These inconsistent empirical results are likely due to the biased evaluation measure used in previous studies. However, it also suggests that more comparisons among different models need to be done to facilitate our understanding in bond rating predictions.

3. Data

3.1. *Sample selection*

The corporate bond data are from the Fixed Income Securities Database (FISD) of the University of Houston. This database includes the month-end data of bonds which compose the Lehman Brother Bond Indices. Many recent papers also use FISD to study the US corporate bond market.[6] The advantage of using FISD is that it differentiates trader quotes from matrix quotes for bond prices. Matrix prices are solely determined by other bonds with the same rating or by adding a fixed spread over Treasury bonds for bonds which are not actually traded in the market. Since bond dealers do not commit to trade these matrix prices, using other popular bond databases which do not make any distinction between trader and matrix prices can be misleading and bias return calculations (Warga and Welch, 1993).[7]

Our bond data cover the period of April 1974 to March 1997. We restrict our sample to industrial bonds only, since we can obtain more bonds from this sector. To reduce the rating prediction biases, we focus on the non-putable senior straight bonds. We also exclude bonds with outstanding amount less than $10 million. To ensure that bond prices are meaningful for long-term investors, we require maturities of five years or longer and trader prices on each April 30 when the rating predictions are performed.

[6]See, for example, Blume, Lim and MacKinlay (1998), Collin-Dufresne, Goldstein and Martin (2001), Eberhart and Siddique (2002) and Maxwell and Stephens (2003).
[7]However, even we include matrix prices in return calculations, the results are qualitatively similar.

Table 1. The rating distribution. This table reports the number of bonds in different ratings and years. Bonds included in the sample must meet the following criteria: (1) industrial bonds, (2) non-putable bonds, (3) senior bonds, (4) straight bonds, (5) with Moody's ratings, (6) with available financial variables listed in the Appendix. Also, on each April 30 when bond ratings are predicted, only bonds with amounts outstanding larger than ten million, time-to-maturities greater than five years, and with trader prices can enter into the sample.

Year	Aaa	Aa	A	Baa	Ba	B	Caa	Ca	Total
74	12	18	21						51
75	20	20	25	4	2	2			73
76	22	26	27	10	2	1			88
77	24	28	27	13	3				95
78	30	65	87	27	5	1			215
79	33	61	89	26	5	3			217
80	36	56	81	24	4	6			207
81	32	54	102	16	9	28	1		242
82	26	38	76	11	6	29			186
83	14	63	91	13	14	29			224
84	13	53	95	19	14	32			226
85	4	60	98	42	25	31			260
86	6	45	132	38	42	28			291
87	11	40	124	44	41	44			304
88	9	34	126	48	23	38			278
89	4	22	112	46	8	20			212
90	5	17	76	27	4	17			146
91	4	18	80	32	7	13	1		155
92	3	19	83	59	10	9	2		185
93	2	21	64	65	27	18	1	1	199
94	3	19	76	56	21	19		1	195
95	4	26	73	64	22	21	1		211
96	4	18	65	68	32	26	1		214
Total	321	821	1830	752	326	415	7	2	4474

Finally, all bonds included in our analysis must have Moody's ratings. The final sample contains 4,474 bond-year observations of 415 firms.

The rating distribution of our sample is reported in Table 1. Since Lehman Brothers was not active in the high-yield bond market until 1992, our sample tilts toward to investment-grade bonds. Less than 18% of our sample is in the speculative grade. The single rating with the most bond issues is A which composes 41% of the sample.

3.2. *Summary statistics*

The summary statistics of our sample are presented in Table 2. As expected, the higher the bond rating, the lower the bond yield. The amounts

Table 2. Summary statistics. For each year from 1974 to 1996, the mean values are first computed within each rating for each variable. The time-series average of these mean values is then calculated. "Years" means the number of years of available data used to compute the time-series average. "Amount" is the outstanding amount of bonds on each April 30, expressed in millions. "Size", in terms of millions, is the market value of equity at previous December end, and "BM" is the book-to-market ratio calculated by dividing book value at previous fiscal year by the market value of equity. "Size Rank" and "BM Rank" are based on Size and BM quintile breakpoints, respectively, obtained from all NYSE stocks. "Stock Return" is the prior six month buy-and-hold return of the corresponding stock.

Ratings	Years	Amount	Yield	Maturity	Size	BM	Size Rank	BM Rank	Stock Return
Aaa	23	225	9.38%	16.920	30547	0.63	5.00	2.30	7.48%
Aa	23	189	9.71%	17.154	11501	0.65	4.99	2.46	11.19%
A	23	139	9.99%	16.095	3970	0.84	4.79	3.10	12.97%
Baa	22	143	10.55%	13.589	1811	0.92	4.26	3.39	16.65%
Ba	22	136	11.60%	12.565	1165	0.89	3.57	3.04	19.42%
B	21	99	14.62%	12.705	286	1.64	2.20	3.42	15.93%
Caa	6	110	18.68%	7.237	87	2.46	1.17	4.83	46.71%
Ca	2	70	N/A	5.541	27	1.86	1.00	3.50	9.09%

outstanding and time-to-maturity generally increase as the rating moves up. In addition, we examine stocks whose corresponding bonds are included in our sample, and get their market value of equity (size) and book-to-market (BM) quintile breakpoints, based on all NYSE stocks on each April 30. As bond ratings move down, the size values decline and BM ratios increase. Since our sample concentrates on investment-grade bonds, the large size rankings for Baa and above indicate that most firms included in this study are big firms. Interestingly, from Aaa down to Ba, the prior six-month buy-and-hold stock returns monotonically increase, but firms with B rating bonds earn less returns than Ba. This suggests that higher default risks do not compensate investors for higher stock returns, a finding consistent with Dichev (1998).

4. Methodology

4.1. *Rating predictions*

We employ four models to predict bond ratings — MDA, MDA with cross-validation (MDA-C), ordered probit (Probit), and ordered probit with stepwise variable selection (Probit-S). The choice of MDA is due to its popularity among previous studies and thus can facilitate the comparison of our results to others.[8] However, MDA requires the multivariate normal distribution for independent variables, and does not take the ordinal nature of bond ratings into account. On the contrary, Probit recognizes that ratings are ordinal and does not require independent variables to be multivariate normal, two theoretical advantages over MDA. By conducting MDA and Probit on the same sample, we can empirically test whether Probit performs better in bond classifications.

Moreover, to improve the performance of MDA, we apply the cross-validation holdout procedure to MDA as the MDA-C model. The cross-validation method is performed by omitting one observation at a time, calculating the classification rule based on remaining observations, and then classifying the holdout observation which is omitted in the beginning. Repeat the above step until all observations are classified. As pointed out in Pinches (1980), this method generates unbiased estimates of error rates, and is reasonably robust to extreme numbers of variables. Similarly, to avoid that large number of independent variables seriously affects the probit regressions, we carry on the stepwise variable selection procedure which picks up a subset of variables that have statistically significant contributions to the model. Adding this procedure on Probit (the Probit-S model) ensures that only the most significant variables are included in the model.

On each April 30 from 1974 to 1996, we perform the rating prediction for all available bonds, by four abovementioned statistical models. The financial variables used to predict bond ratings are selected based on prior studies and listed in the Appendix.[9] Since the rating process by Moody's is conservative and long, it is very likely that Moody's ratings do not capture the true ratings for some periods of time. Accordingly, we assume that the market can have its own assessed ratings, which may be different from ratings supplied by Moody's. For any bond with Moody's rating lower than the predicted rating implied by the model, we will classify it as an under-rated bond since the current rating is worse than it's supposed to be. Similarly, a bond with Moody's rating higher than the predicted rating is grouped as an over-rated bond.

[8]We do not use the OLS model to predict ratings in our results since the accuracy rate of rating predictions shown in the previous research is generally lower for OLS than for MDA.

[9]We assume a four-month reporting lag to collect financial data.

4.2. *Relative rating strength portfolios*

We form a zero-investment portfolio by buying under-rated bonds and selling over-rated bonds at each April 30 when the rating predictions are performed. This zero-investment portfolio is similar to the relative strength portfolio introduced in Jegadeesh and Titman (1993), so we call it the relative rating strength portfolio. Since the previous research suggests that there exist abnormal returns starting from one year and a half prior to rating change announcements (Weistein, 1977) and that the impact of rating revisions can last over three years (Dichev and Piotroski, 2001), we hold the relative rating strength portfolio for five years to track its long-run performance. Examining such long-horizon performance also ensures that our portfolio returns fully capture the market reactions to rating changes, even when rating agencies significantly delay the rating revisions.

To get the bond portfolio returns, we first compute annual buy-and-hold returns for individual bonds by compounding monthly returns from May up to next April, or the final month of the data, whichever is earlier. Equal-weighted portfolio returns are then calculated with an annual portfolio rebalancing for both under- and over-rated bonds.[10] We then obtain the multi-year long-run returns by compounding portfolio annual returns. The relative rating strength portfolio returns are the differences of long-run compounded returns between the under- and over-rated bond portfolios.

We also compute abnormal returns of relative rating strength portfolios by carefully controlling for the bond risks. Specifically, we calculate the market buy-and-hold returns matched by the rating and maturity for each misclassified bond and use them as benchmark returns.[11] Similar to the return calculation mentioned above, annual and long-run compounded market returns are computed. For each of under- and over-rated bond portfolios, long-run abnormal returns are then obtained by subtracting the long-run compounded market returns from those of the bond portfolio. Again the returns of interests are the differences of long-run abnormal returns between the under- and over-rated bond portfolios.

[10]To avoid that extreme returns generated by small number of bonds in one year bias our results, we also calculate the number-weighted average return by multiplying the number of bonds by the relative rating strength portfolio return, summing across years, and then dividing by total number of bonds in the sample period. Though not reported here, the results are very similar to what we report in Table 4.

[11]We are careful in matching the holding horizons between the bond portfolio and the benchmark. For example, if a bond is removed from the return calculation because either it does not have trader price or it is delisted, the market return calculation also stops.

To compute market returns, we use Lehman Brothers corporate bond indices which include all non-convertible bonds with at least one year to maturity and an outstanding amount of $50 million. These indices are categorized by different ratings and sectors, but not jointly. For each rating or sector, there are two indices, intermediate index which is made of bonds with maturities of up to ten years, and long-term index which includes bonds with maturities of ten years or longer. Since we believe that default risks are more important than sectors to bond prices, we choose to control for ratings and maturities, rather than sectors and maturities.[12]

5. Empirical Results

5.1. *Traditional evaluation measure for models to predict ratings*

Table 3 reports the prediction results based on the traditional evaluation measure for different models. The accuracy rate for MDA is about 86% which is higher than that in previous studies. With the 75% accuracy, MDA-C also performs well and dominates similar models in earlier research. Since error rates of MDA-C are proved to be unbiased and relatively robust (Pinches, 1980), this result indicates that at least three-fourths of the Moody's ratings can be replicated by our models. This superior performance is probably attributed to the fact that we include more variables in our models and exclude the subordinated bonds from the sample. Besides, we perform the rating prediction in a year-by-year basis, rather than pool many years together in the training sample as previous studies do, and thus make good use of all available information.

Although Probit has theoretical advantages in predicting bond ratings, surprisingly, its accuracy rate is only 79% which is lower than that of MDA. The Probit-S model, which selects the most significant variables and is expected to be more powerful, works worse than MDA, and even than Probit. This implies that theoretical advantages of the ordered probit model may be not so important empirically. It is also possible that the multiple discriminant analysis model has more power in classifying bonds when the sample comprises a wide range of rating classes.

[12]We take care of rating or maturity changes when compute market index returns as the benchmark. For example, if the bond is downgraded or upgraded to another rating, or its maturity is reducing from long-term to intermediate-term, the benchmark return is carefully matched at the same time.

Table 3. Accuracy rates of bond rating predictions. On each April 30 from 1974 to 1996, the bonds' ratings are predicted by four models: multiple discriminant analysis (MDA), multiple discriminant analysis with cross-validation (MDA-C), ordered probit (Probit), and ordered probit with stepwise variable selection (Probit-S). The variables used to predict ratings are listed in the Appendix. The "No" column reports the total number of bonds used in rating prediction in each year. The "U" column includes the number of under-rated bonds whose Moody's ratings are lower than predicted ratings, while the "O" column contains the number of over-rated bonds whose Moody's ratings are higher than predicted ratings. "Mis" column is the number of misclassified bonds, which is the sum of numbers in "U" and "O" columns. "Accuracy" is the ratio of number of bonds, whose predicted ratings equal to Moody's rating, to total number of bonds in that year.

		MDA				MDA-C				Probit				Probit-S			
Yr	No	U	O	Mis	Accuracy	U	O	Mis	Accuracy	U	O	Mis	Accuracy	U	O	Mis	Accuracy
74	51	0	0	0	100.00%	7	3	10	80.39%	0	0	0	100.00%	4	2	6	88.24%
75	73	1	1	2	97.26%	9	8	17	76.71%	1	2	3	95.89%	3	6	9	87.67%
76	88	1	2	3	96.59%	5	7	12	86.36%	1	2	3	96.59%	6	6	12	86.36%
77	95	2	3	5	94.74%	11	8	19	80.00%	1	2	3	96.84%	7	11	18	81.05%
78	215	13	20	33	84.65%	31	27	58	73.02%	24	16	40	81.40%	26	22	48	77.67%
79	217	13	12	25	88.48%	32	21	53	75.58%	25	18	43	80.18%	29	20	49	77.42%
80	207	15	13	28	86.47%	28	20	48	76.81%	23	16	39	81.16%	26	20	46	77.78%
81	242	17	17	34	85.95%	33	26	59	75.62%	33	20	53	78.10%	25	23	48	80.17%
82	186	8	3	11	94.09%	17	8	25	86.56%	11	15	26	86.02%	13	15	28	84.95%
83	224	12	24	36	83.93%	24	32	56	75.00%	12	16	28	87.50%	18	27	45	79.91%
84	226	12	20	32	85.84%	20	34	54	76.11%	20	24	44	80.53%	25	31	56	75.22%
85	260	19	26	45	82.69%	33	36	69	73.46%	36	37	73	71.92%	43	44	87	66.54%
86	291	32	18	50	82.82%	44	26	70	75.95%	28	36	64	78.01%	31	48	79	72.85%
87	304	33	37	70	76.97%	46	45	91	70.07%	44	41	85	72.04%	39	43	82	73.03%
88	278	20	25	45	83.81%	39	40	79	71.58%	44	32	76	72.66%	45	33	78	71.94%
89	212	15	12	27	87.26%	30	16	46	78.30%	27	17	44	79.25%	21	19	40	81.13%
90	146	7	4	11	92.47%	11	11	22	84.93%	11	5	16	89.04%	16	16	32	78.08%
91	155	11	8	19	87.74%	19	20	39	74.84%	14	16	30	80.65%	29	23	52	66.45%
92	185	12	7	19	89.73%	21	15	36	80.54%	15	11	26	85.95%	20	28	48	74.05%
93	199	23	12	35	82.41%	41	23	64	67.84%	28	32	60	69.85%	32	30	62	68.84%
94	195	17	12	29	85.13%	34	18	52	73.33%	23	29	52	73.33%	27	25	52	73.33%
95	211	26	17	43	79.62%	39	37	76	63.98%	29	28	57	72.99%	36	30	66	68.72%
96	214	19	14	33	84.58%	38	25	63	70.56%	38	31	69	67.76%	44	32	76	64.49%
All	4474	328	307	635	85.81%	612	506	1118	75.01%	488	446	934	79.12%	565	554	1119	74.99%

5.2. *Market-based assessment for models to predict ratings*

In the previous section, we implicitly assume that Moody's ratings are correct in a timely fashion when we employ the traditional measure to evaluate the performance of rating prediction models. In this section, we relax the assumption that agency ratings are always accurate. Table 4 shows the long-run performance of relative rating strength portfolios, formed by buying under-rated bonds and selling over-rated bonds.[13] For raw returns, all models can generate significant returns in the long-run. In particular, over five post-formation years, the relative rating strength portfolio can earn an average return of 11.89% from MDA, 9.9% from MDA-C, 9.57% from Probit, and 8.35% from Probit-S, all of which are significant within 1% level. However, after control for the bond market returns, only MDA-C and Probit-S can generate significant abnormal returns in the long-run. The relative rating strength portfolio based on MDA-C even outperforms that from Probit-S starting from year 2 and generates an abnormal return of 5% over five years.

To check the robustness of our results, we delete below-investment bonds when we form relative rating strength portfolios, and report the result in Table 5. The results are qualitatively similar to Table 4. MDA-C is still the best model among the four models we examine and generate statistically significant abnormal returns in the long-run. However, the abnormal returns from Probit-S is not significant anymore. As a result, there is no evidence that ordered probit models outperform multiple discriminant analysis models in classifying bonds.

There are three implications based on our bond portfolio results. First, MDA-C works the best among four models to predict bond ratings. This indicates that ordered probit models are not superior to multiple discriminant analysis models in classifying bonds even though ordered probit models are more theoretically appropriate. This result is consistent with the finding in Table 3 when we apply the traditional evaluation measure. Second, since MDA-C can identify relative rating strength portfolios which statistically beat the market indices over the long-run but MDA cannot, it suggests that adopting the holdout procedure adds value in rating predictions. The stepwise variable selection also contributes to the rating predictions given that the relative rating strength portfolios from Probit-S outperform those from Probit. Using the traditional evaluation criterion, we observe the

[13]For MDA and Probit, since there are only few misclassified bonds before 1977, we calculate the portfolio return starting from 1978.

Table 4. Long-run bond performance of relative rating strength portfolios. On each April 30 from 1974 to 1996, the bond ratings are predicted by four models: multiple discriminant analysis (MDA), multiple discriminant analysis with cross-validation (MDA-C), ordered probit (Probit), and ordered probit with stepwise variable selection (Probit-S). An under-rated bond portfolio, composed of bonds with Moody's ratings lower than predicted ratings, and an over-rated bond portfolio, composed of bonds with Moody's ratings higher than predicted ratings, are formed and held for five years. The portfolio returns are rebalanced annually for five years and compounded starting from year two. The abnormal returns are computed by subtracting the compounded returns of the corresponding Lehman Brothers bond index from the portfolio compounded returns. The raw (abnormal) returns shown in the row of "mean" are the difference of the compounded (abnormal) returns between under- and over-rated bond portfolios. The average annual number of bonds for under- and over-rated bond portfolios are shown in the rows of "U #" and "O #", respectively. "p-value" is the significance level of the two-side t-test. The numbers in the parenthesis are total numbers of misclassified bonds across years used to compute raw and abnormal returns, respectively, for different models.

		Raw Returns					Abnormal Returns				
Model	Statistics	Year 1	Year 2	Year 3	Year 4	Year 5	Year 1	Year 2	Year 3	Year 4	Year 5
MDA	U #	16	13	11	9	7	16	13	10	8	6
(595,580)	O #	15	14	12	10	7	15	13	11	9	7
	mean	0.77%	2.97%	5.95%	9.10%	11.89%	−0.16%	−0.08%	1.63%	2.20%	3.64%
	p-value	0.161	0.002	0.001	0.001	0.003	0.643	0.903	0.321	0.332	0.219
MDA-C	U #	25	20	17	14	11	23	19	15	12	10
(1057,1005)	O #	21	19	16	13	10	21	19	16	13	10
	mean	1.54%	2.80%	5.59%	8.27%	9.90%	0.51%	1.04%	1.97%	3.25%	4.80%
	p-value	0.020	0.000	0.000	0.000	0.000	0.159	0.081	0.038	0.023	0.016
Probit	U #	24	20	16	14	13	23	19	15	13	12
(882,856)	O #	22	19	16	13	10	22	19	16	13	10
	mean	0.74%	1.96%	1.76%	1.97%	9.57%	−0.29%	−0.58%	−2.09%	−1.56%	−1.01%
	p-value	0.068	0.009	0.055	0.446	0.000	0.374	0.425	0.086	0.350	0.670
Probit-S	U #	23	19	16	13	12	22	18	15	12	10
(1065,1029)	O #	23	20	17	14	10	23	20	16	13	10
	mean	1.18%	1.92%	2.74%	3.52%	8.35%	0.54%	0.47%	0.37%	1.79%	3.57%
	p-value	0.052	0.010	0.002	0.002	0.003	0.164	0.534	0.695	0.181	0.073

Table 5. Long-run bond performance of relative rating strength portfolios: delete below-investment grade bonds. This table is similar to Table 4. The only difference is that below-investment grade bonds are deleted when under- and over-rated bond portfolios are formed on each April 30. The raw (abnormal) returns shown in the row of "mean" are the difference of the compounded (abnormal) returns between under- and over-rated bond portfolios. The average annual number of bonds for under- and over-rated bond portfolios are shown in the rows of "U #" and "O #", respectively. "p-value" is the significance level of the two-side t-test. The numbers in the parenthesis are total numbers of misclassified bonds across years used to compute raw and abnormal returns, respectively, for different models.

		Raw Returns					Abnormal Returns				
Model	Statistics	Year 1	Year 2	Year 3	Year 4	Year 5	Year 1	Year 2	Year 3	Year 4	Year 5
MDA	U #	11	9	7	6	5	11	9	7	6	5
(471,471)	O #	14	13	11	9	7	14	13	11	9	7
	mean	0.69%	1.53%	3.27%	7.65%	11.01%	−0.03%	−0.81%	0.81%	1.66%	2.50%
	p-value	0.142	0.087	0.013	0.042	0.059	0.925	0.351	0.574	0.286	0.214
MDA-C	U #	15	13	10	9	8	15	13	10	9	8
(799,799)	O #	19	18	15	12	10	19	18	15	12	10
	mean	0.95%	1.50%	2.96%	5.11%	6.95%	0.31%	0.35%	0.64%	1.96%	3.85%
	p-value	0.066	0.009	0.003	0.005	0.001	0.283	0.556	0.372	0.073	0.026
Probit	U #	16	14	11	10	10	16	14	11	10	9
(681,681)	O #	20	17	15	12	9	20	17	15	12	9
	mean	0.27%	0.96%	1.32%	2.99%	5.90%	−0.47%	−0.70%	−1.77%	−1.41%	0.28%
	p-value	0.487	0.062	0.130	0.027	0.002	0.232	0.186	0.135	0.407	0.912
Probit-S	U #	16	14	12	10	9	16	14	12	10	9
(848,848)	O #	21	18	16	13	10	21	18	16	13	10
	mean	0.92%	1.16%	1.73%	3.94%	6.98%	0.21%	0.06%	−0.13%	0.70%	2.74%
	p-value	0.112	0.085	0.104	0.086	0.025	0.557	0.926	0.888	0.676	0.212

opposite finding (recall that MDA outperforms MDA-C and Probit beats Probit-S in Table 3). Because the return results here are based on actual traded data after ratings are predicted, our proposed evaluation criterion includes post market information and should be more reliable than the traditional measure. As a result, the different conclusions reached from Tables 3 and 4 suggest that adopting the traditional measure in evaluating statistical models to predict ratings can be misleading. Finally, the major participants of the corporate bond market are institutional investors who are expected to be sophisticated and informed. Therefore, the corporate bond market should be quite efficient. However, the existence of a profitable trading strategy based on MDA-C suggests that the corporate bond market is not so efficient as we expected. One possible reason is that investors rely too much on agency ratings, and the revision process of these agency ratings is too long.

6. Conclusion

Numerous studies have proposed different statistical models to predict corporate bond ratings. These papers use agency ratings as the benchmark to evaluate various prediction models and thus implicitly assume that agency ratings are accurate in a timely fashion. However, the empirical evidence of market reactions to bond rating changes suggests that rating agencies do not adjust bond ratings efficiently to fully reflect the true default risks of bonds. Accordingly, the assumption in previous rating prediction papers is not valid.

Assuming that the market can form an independent assessment about default risks of bonds, we design a new approach to evaluate models in predicting true bond ratings. Specifically, we construct a relative rating strength portfolio formed by buying under-rated bonds, whose Moody's ratings are lower than predicted ratings from the statistical model, and selling over-rated bonds, whose Moody's ratings are higher than predicted ratings. We hold this portfolio for five years after the rating predictions, and examine its long-run performance. If one statistical model can identify misclassified bonds which generate significant higher returns than bond portfolios based on other models, this model has more power to detect bonds whose agency ratings deviate from true ratings, and is a better model in predicting bond ratings.

The evaluation criterion we propose in this paper is superior to that used in the previous research because we include ex-post bond returns into the evaluation and do not assume that ratings from rating agencies are always correct. Moreover, if there does exist a relative rating strength portfolio which earn abnormal returns in the long-run, investors can make profits in the corporate bond market by implementing this trading strategy.

Using 4,474 industrial bonds with Moody's ratings of 415 firms, we perform rating predictions by four statistical models, namely, multiple discriminant analysis (MDA), multiple discriminant analysis with the cross-validation holdout procedure (MDA-C), ordered probit (Probit), and ordered probit with the stepwise variable selection (Probit-S). Based on the traditional evaluation measure, our empirical results show that 86% of agencies' ratings can be correctly replicated by MDA, 75% by MDA-C, 79% by Probit, and 75% by Probit-S. Consistent with previous research, our statistical models fairly capture the essence of the rating process of rating agencies. However, since Probit and Probit-S do not perform better than MDA or MDA-C, it implies that theoretical advantages of ordered probit models do not guarantee the superior performance.

With our alternative approach, we find that only MDA-C can identify bond portfolios which consistently generate significant abnormal returns in the long-run (about 5% over a five-year horizon). This indicates that the evaluation criterion used in previous studies may be misleading since MDA-C performs better than MDA in our new approach but not in the traditional one. Moreover, to the extent that we document a profitable trading strategy through MDA-C, our result suggests that the corporate bond market is not very efficient even though most of the market participants are sophisticated institutional investors.

Appendix: Variables Used in the Statistical Models

A.1. Profitability

1. Pretax return on permanent capital = EBIT/average total assets
2. Return on asset = net income/total assets
3. Return on stock = last 12 month buy-and-hold stock return ending on previous December
4. Return on stock = last six month buy-and-hold stock return ending on previous December
5. E/P ratio = earnings per share/price
6. D/P ratio = dividend per share/price

A.2. Earnings variability

1. ROA variability = standard deviation of recent five year returns on assets
2. Pretax return variability = standard deviation of recent five year EBIT to total assets ratios

A.3. Coverage

1. Pretax interest coverage = EBT/interests
2. Pretax interest coverage including rent = (EBT + rent)/(interests + rent)
3. EBITDA interest coverage = (EBT + rent + depreciation)/interests

A.4. Cash flow to debt ratio

1. Cash flow/total debt = (net income + depreciation)/(long-term debt + short-term debt)
2. Cash flow/long-term debt = (net income + depreciation)/long-term debt
3. Free cash flow/total debt = (net income + depreciation − capital expenditure)/(long-term debt + short-term debt)
4. Free cash flow/long-term debt = (net income + depreciation − capital expenditure)/long-term debt

A.5. Leverage

1. Long-term debt to capitalization = long-term debt/total assets
2. Short-term debt to long-term debt
3. Deferred taxes to long-term debt

A.6. Firm size

(Based on previous December price and number of shares in CRSP; all data but price are in millions before taking logarithm)

1. Market value of total asset = log (price * number of shares + total assets − book value of common equity)
2. Market value of common equity = log (price * number of shares)

A.7. Growth potential

1. BM ratio = book value of common equity/market value of common equity
2. Sales growth in recent three years = log (sales in year $(t-1)$/sales in year $(t-3)$)

A.8. Operating efficiency

1. Asset turnover = sales/average total assets
2. Receivables turnover = sales/average receivables

A.9. Liquidity

1. Current ratio = current assets/current liabilities
2. Quick ratio = (current assets − inventory)/current liabilities

A.10. Other

1. Size of bonds = log (outstanding amount of the bond in thousands on April 30)
2. Dividend to interest ratio

References

Blume, M, F Lim and C MacKinlay (1998). The declining credit quality of US corporate debt: Myth or reality? *Journal of Finance*, 53, 1389–1413.

Collin-Dufresne, P, R Goldstein and S Martin (2001). The determinants of credit spread changes. *Journal of Finance*, 56, 2177–2207.

Dichev, I (1998). Is the risk of bankruptcy a systematic risk? *Journal of Finance*, 53, 1131–1147 .

Dichev, I and J Piotroski (2001). The long-run stock returns following bond ratings changes. *Journal of Finance*, 56, 173–203.

Eberhart, A and A Siddique (2002). The long-term performance of corporate bond (and stocks) following seasoned equity offerings. *Review of Financial Studies*, 15, 1385–1406.

Ederington, L (1985). Classification models and bond ratings. *The Financial Review*, 20, 237–262 .

Ederington, L and J Yawitz (1987). The bond rating process. In *Handbook of Financial Markets*, E Altman (ed.). New York: John Wiley & Sons.

Eisenbeis, R (1977). Pitfalls in the application of discriminant analysis in business, economics, and finance. *Journal of Finance*, 32, 875–900.

Fama, E and K French (1992). The cross section of expected stock returns. *Journal of Finance*, 47, 427–465.

Hite, G and A Warga (1997). The effect of bond-rating changes on bond price performance. *Financial Analysts Journal*, 53, 35–51.

Holthausen, R and R Leftwich (1986). The effect of bond rating changes on common stock prices. *Journal of Financial Economics*, 17, 57–89.

Horrigan, J (1966). The determination of long-term credit standing with financial ratios. *Journal of Accounting Research*, 4, 44–62.

Howe, J (1995). Credit analysis for corporate bonds. In *The Handbook of Fixed Income Securities*, F Fabozzi and D Fabozzi (eds.). Chicago, IL: Irwin, Inc.

Jegadeesh, N and S Titman (1993). Returns to buying winners and selling losers: Implication for stock market efficiency. *Journal of Finance*, 48, 65–91.

Johnson, R and D Wichern (1992). *Applied Multivariate Statistical Analysis*, 3rd Ed. Englewood Cliffs, NJ: Prentice-Hall.

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

Lachenbrush, P, C Sneeringer and L Revo (1974). Robustness of the linear and quadratic discriminant function to certain types of non-normality. *Communications in Statistics*, 1, 39–57.

Maxwell, W and C Stephens (2003). The wealth effects of repurchases on bondholders. *Journal of Finance*, 58, 895–919.

McKelvey, R and W Zavoina (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, 4, 103–120.

Pinches, G (1978). A multivariate analysis of industrial bond ratings and the role of subordination: Reply. *Journal of Finance*, 32, 336–344.

Pinches, G (1980). Factors influencing classification results from multiple discriminant analysis. *Journal of Business Research*, 8, 429–456.

Pinches, G and K Mingo (1973). A multivariate analysis of industrial bond ratings. *Journal of Finance*, 28, 1–18.

Pinches, G and K Mingo (1975). The role of subordination and industrial bond ratings. *Journal of Finance*, 30, 201–206.

Pinches, G and JC Singleton (1978). The adjustment of stock prices to bond rating changes. *Journal of Finance*, 33, 29–44.

Standard & Poor's (1998). *Corporate Ratings Criteria*. New York: McGraw-Hill.

Warga, A and I Welch (1993). Bondholder losses in leveraged buyouts. *Review of Financial Studies*, 6, 959–982.

Weinstein, M (1977). The effect of a rating change announcement on bond price. *Journal of Financial Economics*, 5, 329–350.

Wingler, T and J Watts (1982). Electric utility bond rating changes: Methodological issues and evidence. *Journal of Financial Research*, 5, 221–235.