韓國經營科學會誌
第26卷 第1號
2001年 3月

AR 프로세스를 이용한 도산예측모형*

이군희** · 지용희**

Bankruptcy Prediction Model with AR process*

Gun-Hee Lee** · Yong-Hee Chee**

■ Abstract ■

The detection of corporate failures is a subject that has been particularly amenable to cross-sectional financial ratio analysis. In most of firms, however, the financial data are available over past years. Because of this, a model utilizing these longitudinal data could provide useful information on the prediction of bankruptcy. To correctly reflect the longitudinal and firm-specific data, the generalized linear model with assuming the first order AR (autoregressive) process is proposed. The method is motivated by the clinical research that several characteristics are measured repeatedly from individual over the time. The model is compared with several other predictive models to evaluate the performance. By using the financial data from manufacturing corporations in the Korea Stock Exchange (KSE) list, we will discuss some experiences learned from the procedure of sampling scheme, variable transformation, imputation, variable selection, and model evaluation. Finally, implications of the model with repeated measurement and future direction of research will be discussed

Keyword : bankruptcy prediction model, generalized linear models, AR process, longitudinal data analysis, analysis with repeated measurements, Generalized Estimating Equations (GEE), logit link function

논문접수일 : 2000년 11월 23일 논문게재확정일 : 2001년 3월 5일
* This research was supported by the Sogang University Research Grants in 2000 and Sogang Institute for Entrepreneurial Studies.
** College of Business Administration, Sogang University

1. Introduction

The detection of corporate failures is a subject that has been particularly amenable to cross-sectional financial ratio analysis. Most of bankruptcy models have proposed one-period prediction model (e.g., Beaver, 1966 ; Altman, 1968 ; Ohlson, 1980 ; Zavgren, 1985). With the time-dependent predictors, there are two aspects about how to combine the prediction model with time. The first aspect is concerning as to how to set the origin of time. If the origin of time is set as the year of bankruptcy, the predictors with same year prior to bankruptcy have different calendar-year. However a typical one-period prediction model ignores these changes. The second aspect is that one-period prediction model uses only one set of predictors even the financial data are available over past years. In this paper, Several ways to overcome these difficulties are presented by introducing the two different times, calendar-year and years prior to bankruptcy.

The remainder of the paper is structured as follows : The next section introduces the model with repeated measurement. In section three, research design and data are presented followed by analysis. Both the benefits and limitations of this new approach as well as future research directions will be discussed later.

2. The Model

The model presented here extends one-period prediction models in two ways : (1) allowing repeated indicators over years prior to bankruptcy ; and (2) allowing financial ratios as time-dependent predictors. The first extension investigates any patterns of predictors when a firm is close to bankruptcy. In this case, the origin of time is set to the year in bankruptcy. For non-bankrupt firms, the origin of time is set to be the year of study. The second extension enables the predictors to compare over different calendar-year by handling the yearly effect as the block effect in the randomized block design. For simplicity, we assume throughout the paper that bankruptcy can occur at discrete points in time.

Let $X_{ijtt'}$ be the i-th firm and j-th financial ratio at time t and t', where t and t' be the year prior to bankruptcy and corresponding calendar-year, respectively. In generalized linear model(Nelder and Wedderburn, 1972), the firms are viewed as a random sample from a population. With J predictors, the generalized linear model is the following additive model:

$$g(p_{it}) = \mu + \gamma_{t'} + f_i + d_{t(i)} + \beta_1 X_{i1tt'} + \beta_2 X_{i2tt'} + \cdots + \beta_J X_{iJtt'},$$

where $g(\cdot)$ is a link function, μ is a intercept, $\gamma_{t'}$ is a calendar-year effect, f_i is a firm-specific random effect with assuming independent random variable with mean 0 and firm-specific heterogeneous variance σ_i^2, $d_{t(i)}$ is the effect of t-th year prior to bankruptcy within i-th firm and β_j is a regression coefficient of j-th predictor. Note that if we assume the binomial-type likelihood function, the firm-specific heterogeneous variance, σ_i^2 is proportional to $p_i(1-p_i)$, where p_i is the probability of bankruptcy for i-th firm. In this case, the likelihood function of the first stage model becomes

$$L(p_{it}) \propto p_{it}^{y_{it}}(1-p_{it})^{(1-y_{it})},$$

where y_{it} is 1 if i-th firm is bankrupt and 0 otherwise. It is reasonable to assume that there is a correlation between p_{it}'s with same i-th firm. It is thus further assumed that the correlation structure follows the first-order autoregressive(AR) process:

$$Corr(p_{it}, p_{i(t-1)}) = \rho_i,$$
$$Corr(p_{it}, p_{i(t-2)}) = \rho_i^2,$$
$$\vdots$$
$$Corr(p_{it}, p_{i(t-k)}) = \rho_i^k.$$

The statistical inference of correlated data arising from repeated measurement with categorical response can be accomplished by Generalized Estimating Equations (GEEs) method (Liang and Zeger, 1986). If we choose a link function, g_{it} as logit (p_{it}), the second stage model becomes logistic regression model. In this study, two hierarchical models are used : (1) logit link and binomial-type likelihood function with autoregressive correlation structure (refer to as the Binomial - Logit - AR model) and (2) log link and poisson-type likelihood function with unstructured correlation (refer to as the Poisson - Log - Unstructured model).

Another methodology to handle the repeated measurement is summary statistics approach (Diggle, Liang and Zeger, 1994). To eliminate the yearly effect, normal scores, Z_{ijt}'s are used based on $X_{ijtt'}$ for subpopulation at t' point. This transformation eliminate not only the yearly effect but the leverage effect which might contaminate the prediction model. It is further assumed the following firm-specific model for each i.

$$Z_{ijt} = \alpha_{ij} + \beta_{ij}t + \varepsilon_{ijt},$$

where ε_{ijt} represents normally distributed error with assuming mean 0 and common variance σ^2 and β_{ij} represents average rate of change for j-th financial ratio of i-th firm. If the origin of time is set as the year of bankruptcy and select a specific years prior to bankruptcy (e.g. the 5-th year prior to bankruptcy) the intercept, α_{ij} might be considered as a normal score related with j-th financial ratio which does not influenced by bankrupt conditions. Therefore, the average of α_{ij} is expected to be 0. This fact suggests the regression model without intercept, that is

$$Z_{ijt} = 0 + \beta_{ij}t + \varepsilon_{ijt}.$$

In this case, β_{ij} represents the average rate of Z_{ijt} when t changes. When $t=0$, j-th financial ratio is set to be the average of the financial ratio. Finally, with the additivity on the logit scale, the model becomes the following logistic regression model :

$$\text{logit}\,(p_i) = \gamma_0 + \gamma_1\beta_{i1} + \cdots + \gamma_j\beta_{ij},$$

where γ_j represents regression coefficient for j-th financial ratio. The derived variable, β_{ij} can be used to apply any other predictive model such as tree based model and neural network model.

3. Research Design and Data

Korea was selected as the sample frame. Unlike other countries, Korea was stable throughout most of the 1990's and faced to economic crisis officially appeared in the 4th quarter of 1997 (even though there had been early warning signs since the 2nd quarter). Because of the crisis, an International Monetary Funds (IMF) bailout package of standby credit was delivered in mid-December, albeit with harsh mandatory IMF restructuring agreements placed into effect. Since then, an unprecedented number of corporations have gone bankrupt, because banks were reluctant to loan money, causing firms to suffer from high interest rates, loans from abroad were similarly blocked, and strong labor union hindered necessary layoffs. Because of these, the economic environment may change (Sung, Chang and Lee, 1999). All bankrupt corporations between the 2nd quarter of 1997 and the 3rd quarter of 1998 are identified from the Korea Stock Exchange (KSE) list.

3.1 Data Collection

Bankrupt firms were referred to as an act of filing a petition for bankruptcy reported by the KSE. However, both "chaebo" companies and small-sized firms were excluded from the sample. It was because chaebol companies showed totally obscure financial structures (due to mutual loan guarantees and manipulated financial statements) and also it was almost impossible to collect financial statements for small firms. Surprisingly, most of the bankrupt corporations were in manufacturing industries, with only 2 firms in service, 1 in financing, and 2 in the construction industries. Because of such uneven bankrupt distribution across industries, the prediction model can be influenced by undesirable effect. Thus, 70 manufacturing firms were included in bankrupt group. As the control sample, Non-bankrupt group consists of 393 manufacturing firms out of 413 firms listed on the KSE. After excluding 20 firms due to inappropriate results of auditing, all financial data was gathered directly from the KSE starting from the 1st quarter of 1993 for bankrupt and non--bankrupt firms. In particular, for bankrupt firms, the most recent financial statement reported on 6 months prior to the bankrupt date was treated as the first year prior to bankruptcy. Finally, all financial ratios were transformed to the normal score based on each calendar year.

3.2 Variable Selection

Through extensive literature review on bankruptcy prediction, 52 financial ratios categorized as growth, profitability, safety / leverage, activity / efficiency, and market value / dividend are investigated. Out of a total 52 of financial ratios (variables), several variables were found to be significant indicators of bankruptcy. Finally, 8 financial ratios were selected after careful analysis of characteristics of each financial ratios : growth rate of sales (GRS), dividends to capital stock (DCS), quality of operation earnings (QOE), stockholders' equity to total asset(SEA), cash flow to liabilities (CFL), sales to total assets (SLA), inventory turnover (IVT) and market value (MVE).

4. Analysis

The performance of model depends on the misclassification cost as well as prior probability setting. Suppose that the model classifies all firms as non-bankrupt firm. In this case the sensitivity equals to 0, the specificity equals to 1 and total accuracy equals 393 / (393 + 70) = 0.849. Clearly, this model should not be used although it has high accuracy. Therefore in our study, rather than using accuracy, equal cost for each bankrupt and non-bankrupt group is used with the Bayes rule to classify the firm. In this case, the cuttoff probability becomes as follows :

$$\text{cutoff prob.} = \frac{1}{1+393/70} = 0.151$$

In experimental design terminology, the models presented in section 2 have the calendar-year effect treated as a block. Hence $X_{ijtt'}$ is adjusted by the average of financial ratios in the calendar-year. Normal scores, Z_{ijt}'s are used with $X_{ijtt'}$ for subpopulation at t' point. This transformation eliminates the yearly effect as well as the leverage effect which might affect the prediction model because of influential points or outliers.

4.1 Binomial-Logit-AR Model

To select the appropriate model, the backward elimination method is used. With 0.05 significant level, QOE, DCS, IVT and GRS were removed, sequentially. The final model includes CFL, SLA, SEA and MVE. After 9 iterations with GEE method, <Table 1> displays final parameter estimates with asymptotic significance, Z score and p value. As shown in <Table 1>, the degree of importance is CFL, SEA, MVE, SLA, in orderly manner.

〈Table 1〉 Parameter Estimates of the Binomial-Logit-AR Model

Variable	Estimate	Std. Err.	Z score	p value
SEA	0.0436	0.0104	4.174	0.0000
CFL	0.0170	0.0028	5.987	0.0000
SLA	0.0367	0.0112	3.285	0.0010
MVE	0.0344	0.0098	3.496	0.0005

The average risk is computed using the same rule of misclassification cost described in the previous section. That is, if bankrupt firm is classified as non-bankrupt firm, the risk is 5.6. If non-bankrupt firm is classified as bankrupt firm, the risk becomes 1. The final results are shown at <Table 2>.

〈Table 2〉 The Performance of the Binomial-Logit-AR Model

Years	Sensitivity	Specificity	Accuracy	Ave. Risk
1st	0.7429	0.6513	0.6646	0.5068
2nd	0.6714	0.7094	0.7039	0.5151
3rd	0.5857	0.7337	0.7122	0.5640
4th	0.6286	0.7337	0.7184	0.5292

4.2 Poisson-Log-Unstructured Model

With CFL, SLA, SEA and MVE, the GEE method provides parameter estimates after 50 iterations. <Table 3> shows final parameter estimates with asymptotic significance. It is indicated that the degree of importance is CFL, SLA, SEA, MVE in order. This result is slightly different from that with the binomial-logit-AR model.

〈Table 3〉 Parameter Estimates of the Poisson-SLog-Unstructured Model

Predictor	Estimate	Std. Err.	Z score	p value
SEA	0.2412	0.0698	3.455	0.0006
CFL	0.2654	0.0423	6.269	0.0000
SLA	0.3042	0.0627	4.849	0.0000
MVE	0.1926	0.0776	2.483	0.0130

In the comparison with the binomial-logit-AR model, the poisson-log-unstructured model have always higher specificities and lower sensitivities. It can be conjectured that different specification of models provides different strength and weakness and thus produces different results even though the models are based on the same configurations. The final results are shown at <Table 4>.

〈Table 4〉 The Performance of the Poisson-Log-Unstructured Model

Years	Sensitivity	Specificity	Accuracy	Ave. Risk
1st	0.7000	0.6683	0.6729	0.5271
2nd	0.6571	0.7288	0.7184	0.5101
3rd	0.5429	0.7579	0.7267	0.5781
4th	0.6143	0.7482	0.7288	0.5284

4.3 Derived Variable Analysis

As mentioned in Section 2, the analysis is done using derived variables β_{ij}'s. The stepwise procedure with 0.15 significant level as entry and 0.30 significant level as removal criteria was applied to select variables. Three variables were selected as the result of the stepwise procedure. CFL, SLA, SEA are included in the final model. With these predictors, we apply two prediction models : logistic regression model (LRM) and multivariate discriminant model (MDA). In LRM, the final estimated function is as follows :

$$\begin{aligned}\text{logit}(p_i) &= \log\left(\frac{p_i}{1-p_i}\right)\\ &= 2.26 + 1.18\beta_{i4} + 3.27\beta_{i5} + 1.45\beta_{i6},\end{aligned}$$

where p_i is a probability to bankruptcy, β_{i4}, β_{i5} and β_{i6} are derived variables for SEA, CFL and SLA, respectively. The detailed descriptions of parameters are shown at

<Table 5>. In MDA, the final discriminant function is as follows :

$$Z_i = 0.98\beta_{i4} + 2.62\beta_{i5} + 1.42\beta_{i6}$$

In the comparison of means for three variables, it was observed that bankrupt firms showed lower CFL, SEA and SLA, while non-bankrupt corporations showed otherwise. This result matched the general belief of financial experts, which showed the face validity of the prediction model. The means of discriminant scores for bankrupt and non-bankrupt groups were -0.8920 and 0.1589, respectively. The performance of both LRM and MDA is shown at <Table 6>.

〈Table 5〉 Parameter Estimates in Derived Variable Analysis with LRM

Predictor	Estimate	Std. Err.	Std. Estimate	p value
SEA	1.1804	0.4754	0.2117	0.0130
CFL	3.2677	0.6787	0.4788	0.0001
SLA	1.4498	0.4613	0.2517	0.0017

〈Table 6〉 The Performance in Derived Variable Analysis with LRM, MDA

Model	Sensitivity	Specificity	Accuracy	Ave. Risk
LRM	0.7571	0.6870	0.6976	0.4713
MDA	0.7714	0.6819	0.6955	0.4635
Traditional MDA	0.4921	0.5833	0.5689	0.8011

For comparison purpose, we performed traditional multivariate discriminant analysis (Altman, 1968) with use of only one set of financial data set which is presented one year prior to bankruptcy. It is shown that the traditional MDA provides lower specificity and sensitivity with high average risk.

5. Discussion

In our study, several models utilizing longitudinal financial data are proposed. In the application of bankruptcy prediction model among KSE data base, it is shown that the proposed models are superior to traditional model in a sense of average risk. However, care should be taken for the interpretation of results. First of all, there is no theoretical evidence (such as consistency or unbiasedness) for the estimate of the proposed models. Therefore, theory behind the proposed models is highly depending on the asymptotics and the large sample principle. Therefore, it is hard to say the proposed models are always better than traditional bankruptcy model such as traditional MDA. However, unlike other bankruptcy prediction model, the good advantage of this model is that (1) the models allow several sets of past financial data, (2) control calendar year as well as years prior to bankruptcy and (3) consider several sets of past financial data in the same firm as one block.

Our analysis tries to solve the selection bias problem as mentioned in Zmijewski(1984) by setting misclassification cost as 70 / 393 for sensitivity case and 1 for specificity case. But this setting is quite arbitrary. By careful investigation of different setting of the results, we conclude that the model performance might not change with somewhat different setting of misclassification cost.

One of the critical weakness of our proposed model is that dependent variable of firm status (bankruptcy / non-bankruptcy) is not independent over time. Since financial data are available till bankruptcy, the value point of dependent variable changes the covariate structure as well as previous values of dependent. In other words, if we know a firm is non-bankruptcy status in 1997, without any further information, we already know the firm's status is non-bankruptcy in 1996. This fact implies dependent variables for different year are totally dependent each other. This point is major difference between bankruptcy prediction model and traditional multivariate repeated measurement analysis. However, we would like to put this problem as future study for the longitudinal bankruptcy prediction model.

References

[1] Agresti, A. (1990). *Categorical Data Analysis*, New York : John Wiley and Sons.

[2] Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *Journal of Finance*, **23**, Vol.3, pp.589-609.

[3] Altman, E., Haldeman, R., and Narayanan, P. (1977). ZETA Analysis : A New Model to Identify Bankruptcy Risk of Corporations, Journal of Banking and Finance, **1**, Vol.1, pp.29-54.

[4] Altman, E. (1993). *Corporate Financial Distress and Bankruptcy*. New York : John Wiley and Sons.

[5] Bank of Korea (1998). *Financial Statement Analysis*, Seoul, Korea : Bank of Korea Press.

[6] Beaver, W. (1966). Financial Ratios as Predictors of Failures, *Journal of Accounting Research*, **4**, Vol.1, pp.71-127.

[7] Bishop, Y., Feinberg, S.E. and Holland, P.W. (1975). *Discrete Multivariate Analysis : Theory and Practice*, Cambridge, MA : MIT Press.

[8] Deakin, E. B. (1972). A Discriminant Analysis of Predictors of Business Failure, *Journal of Accounting Research*, **10**, Vol. 1, pp.167-179.

[9] Diggle, P.J., Liang, K.Y. and Zeger, S.L. (1994). *Analysis of Longitudinal Data*, Oxford : Clarendon Press.

[10] Dobson, A. (1990). *An Introduction to Generalized Linear Models*, London : Chapman and Hall.

[11] Korea Industrial Bank (1997). *Financial Management*, Seoul, Korea : Korean Industrial Bank Press.

[12] Lawless, J.E. (1982). *Statistical Models and Methods for Lifetime Data*, New York : John Wiley and Sons.

[13] Liang, K.Y. and Zeger, S.L. (1986). Longitudinal Data Analysis using Generalized Linear Models, *Biometrika*, **73**, pp.13-22.

[14] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, London : Chapman and Hall.

[15] Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized Linear Models, *Journal of the Royal Statistical Society A*, **135**, pp.370-384.

[16] Ohlson, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research*, **18**, Vol.1, pp. 109-131.

[17] SAS Institute Inc. (1997). *SAS/STAT Software : Changes and Enhancements through Release 6.12*, Cary, NC : SAS Institute Inc.

[18] Sung, T., Chang, N. and Lee, G. (1999). Dynamics of Modeling in Data Mining : Interpretive Approach to Bankruptcy Prediction, *Journal of Management Information Systems*, **16**, No.1, pp.63-85.

[19] Zavgren, C. V., Dugan, M. T. and Reeve, J. M. (1988). The Association between Probabilities of Bankruptcy and Market Responses : A Test of Market Anticipation, *Journal of Business Finance and Accounting*, **15**, Vol.1, pp.19-45.

[20] Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models, *Journal of Accounting Research*, **22**, Vol.1, pp.59-82.