Credit Risk Modeling: Default Probabilties

Jaime Frade

Department of Statistics
Florida State Univeristy

Dr. XuFeng Niu
STA5168

December 19, 2008

1. Introduction
   - Objective
   - Preliminaries
     - Financial Definitions
     - Description of Data
     - Data: Predictor Variables
   - Historical Methods of Quantitatively Scoring
2. Logistic Regression Models
   - Process of Determining a Model
   - Saturated Logistic Regression Model
   - Logistic Regression: Model 2
3. Comparsion and Prediction
   - Prediction
   - Comparsion
4. Validation Techniques
5. Conclusion

Credit Risk Methodolgy

In recent years more emphasis has shifted to the modeling and evaluation of credit risk. There are several forces behind this trend. First, credit markets have grown steadily and the credit derivatives market (CDS) has grown exponentially. Second, the understanding and the methodologies in managing and measuring other market risks, including interest rate risk, currency risk etc., have matured and are now assessed based on widely accepted principles, while credit risk is still a developing field.

Goal

Built a quantative model that estimates the probability that a US issuer will default on public debt within a year, as a function of predictors that include financial ratios and equity market variables. The model is structured to balance timeliness and stability, and also allow company- or industry-level qualitative analysis.

The basic output of the model is a forward estimate of default probability. The model is uses logistic regression to derive the risk of insolvency within a year as a function of financial and market ratios.

- Determine a logistic regression model
- After determining a statisitcally significant model, will show how to estimate probabilities from current data and intepret a logit model.
- Discuss the significance and relationship of the variables.
- Comparsion of the model with other currently accepted models.
- Apply validation techniques.

Financial Definitions

When a corporation offers a stock or bond for sale, or a government offers a bond, the security is known as an **issue**, and the company or government is the **issuer.**

- Bonds: A debt instrument issued for a period with the purpose of raising capital by borrowing for the issuer.
- Default: The inability of a borrower to pay the interest or principal on a debt when it is due

Data

Obtianed quarterly financial and market information from different issuers (ranging from publicly-traded US bank holding companies to industrial industries, etc) during 1998-2008.

- Issuers consisted of myriad of companies that have either defaulted or are currently stable
- The data for the predictors was obtained from the financial statements from each issuer.

Defaulted Bonds

Defaulted bonds were chosen industry and time independent. Issuers that defaulted most (but not all) of these subsequently facing regulatory intervention as well.

- **Airlines:** Delta Air Lines Inc, Northwest Airlines Corporation, FLYi, Inc,
- **Financial Insitutions:** Lehman Brothers, Bear Stearns, WAMU
- Delphi Corporation, Levitz Home Furnishings, Inc, Winn-Dixie Stores Inc, Calpine Generating Co LLC

Data: Predictors should target explaining/correlating with the following framework

The most commonly used typology of risk factors for issuers is the one regulators uses the CAMEL framework. The focus is on determining predictors that targeted the following general areas which affect default concerns.

- Capital Adequacy: Amount of core capital to a issuer's risk-weighted assets.
- Assets quality.
- Management quality.
- Earnings: most common index of earnings is the ratio of net income to total assets.
- Asset Liquidity.

Data: Predictors for Measuring Insolvency

The predictors will include financial ratios of some of the variables below, in addition to equity-driven measures.

- **Working Capital:** Operating liquidity available to a business, calculated as current assets minus current liabilities
- **Retained Earnings:** The percentage of net earnings not paid out as dividends, but retained by the company to be reinvested in its core business or to pay debt.
- **Earnings Before Debit and Interest (EBIT):** A company's earning power, represents the earnings which the company has achieved. Measure of a issuer's profitability.
- **Market Value of Equity:** The price at which investors buy or sell a issue.

Data: Predictors for Measuring Insolvency (cont.)

- **Equity Market Volatility:** movements in the market in which it is traded
- **Sales:** Total dollar amount collected for goods and services provided
- **Total Liabilities:** The liabilities found by adding current liabilities to long-term debts.
- **Total Assets:** All the property owned by a corporation
- **Net Income:** The company's total earnings, minus expenses and taxes

Existing quantitative models for scoring issuers' credit risk fall into two broad classes: statistical and structural.

- **Statistical Models:** Use historical data on characteristics of issuer (for example, measures of earnings or liquidity) to determine the set of characteristics that best predict the occurrence of the selected outcome. The precise form of the relationship between the inputs and the outcome is specified by the particulars of the statistical model used. These models are historically specific: the model parameters depend on the data used to create the model.
- **Structional Models:** models are form and parameters are, in principle, specified by a theory of economic structure. Contingent claims models, which basis the equity of a firm as a *call option*. The best-known structural models are in the context of the Black-Scholes (1973) framework, which is applied by Merton Approach (1974).

Logistic Regression Model:

I apply a statistical model to historical data on issuers' characteristics. The particular form of statistical model is a discrete-time event history model. This model is designed to predict the risk of an event occuring, as a function of specified variables measured before the event occurs.
The linear regression (a discrete time model) can be used to predict the risk of an event within a certain time period. This is equalvent and estimated by applying a logistic regression to issuer-year of data.
The logistic regression takes the following form

$$\log\left(\frac{p}{1-p}\right) = \sum_{k=1}^{K} \beta_k \, x_k \tag{1}$$

where p is the probability of the even occuring, and K independent variables, x, are each weighted by a coeffiecent, β.

From (1), this different from the conventional linear regression model

$$y = \sum_{k=1}^{K} \beta_k \, x_k \tag{2}$$

From (2) differs from (1), because does not predict the value of the dependent variable, y, but rather the natural logarithm of the ratio of the probability of the event occuring to the probability of the event not occuring.
This logit function preferable to the linear regression function because it limits p, the probability of the event, to be between 0 and 1 (or zero and 100%), where applying the conventional linear regression to dichotomous outcomes would instead allow nonsensical results like probabilities greater than 1 or less than zero.

A transformation of (1), obtain the logistic model which gives the probabilty of the event as

$$p = \frac{e^{\sum_{k=1}^{K} \beta_k x_k}}{1 + e^{\sum_{k=1}^{K} \beta_k x_k}} \tag{3}$$

This difference in structure between this model and more familiar linear regression models gives rise to differences in how the results of the model can be interpreted. Changing one independent variable by a fixed amount changes the level of the dependent variable by an amount that is identical, no matter what the levels of the other independent variables.

If intepreting a linear regression to predict the risk of a bank failing, whether its current risk was 10%, decreasing its ratio of net income to assets by a fixed quantity would increase its risk of failure by the same amountlet's say 1%, respectively 11%.

In contrast, in a logistic regression, a change in one factor changes the risk by an amount that is **proportional** to the level of the other factors. If the other factors produce a failure probability of 2%, decreasing the ratio of net income to assets might double the risk to 4%.

Choosing a Significant Model

Am estimating algorithm is used to find the coefficients, β's that best satisfy the relationship expressed in the regression equation for the estimation data sample. The technique used to find those coefficients for logistic regression, was using maximum likelihood estimation. Basically, method tries coefficients until it finds the set that maximizes the value of a mathematical function that gives the joint probability of observing the given data.

That function, L, the likelihood function, forms the basis of a statistical test of how well the model fits the observed data:

$$L^2 = 2(\log L_T - \log L_B) \tag{4}$$

where L_T is the likelihood function of the first model with smaller variables and the L_B is the liklihood function of a baseline model.

L^2 is a statisic that will be compared with the standard χ^2 table to determine whether the tested model fits significantly better than a baseline model. This procedure is completed to establish whether the added variables significantly improve the fit to the data, or whether conversely the smaller subset is equally sufficient. We use this procedure to omit from the model variables that do not significantly improve our ability to predict insolvency.

- **Working Capital/Total Assets:** Measures the short-term liquidity of the issuer
- **Retained Earnings/Total Assets:** Measures Historic Profitabilty
- **EBIT/Total Assets:** Measures Current Profitabilty
- **Market Equity/Total Liabilities** Measures Leverage
- **Sales/Total Assets:** Approxmated the competitive situation of the company

- **Working Capital/Total Assets:**
- **Retained Earnings/Total Assets:**
- **EBIT/Total Assets:**
- **Market Equity/Total Liabilities**
- **Sales/Total Assets:**

Prediction of Probabilties of Default

Comparsion of Other Quanitative Models and Measurements.

- CreditSights BondScore CDS Model
- Barclay's (Lehman) Corporate Default Model
- Altman Z-Score

Assessment of Final Model

- Cumulative Accuracy Profile (CAP)
- Receiver Operating Profile

References

- yy