

Credit Risk Models I

Week 3

Daniele Bianchi¹

whitesphd.com

¹School of Economics and Finance
Queen Mary, University of London

This week we introduce basic concepts of credit risk models with a particular emphasis on the dichotomy “structural” vs “reduced-form” models. Based on advanced research work we are going to introduce machine learning methods for the analysis of credit risk models.

1. Structural Models of Credit Risk
2. Reduced-form Approach to Credit Risk
3. Applying ML Tools for Default Prediction

Structural Models of Credit Risk

- An important application of big data approaches is in the forecasting of (corporate and retail) default.
- Predicting default useful in many ways:
 - Knowing the risk profile of loans allows you to build an optimal loan portfolio.
 - Loan pricing: $\text{Risk} = \text{Expected return}$.
 - Regulatory requirements: Value at risk computation for a loan portfolio.
 - Distress risk for firms helpful for forecasting equity returns, since distress risk should also be reflected in stock returns.
 - Predicting bond and Credit Default Swap (CDS) prices, etc.

Structural and Reduced Form Approaches

- Two main approaches used to predict (corporate) default (with analogies in individual default predictions):
 1. Structural approaches to default forecasting.
 2. Reduced-form approaches to default forecasting.
- **Structural Approach:** There is an underlying economic model. Imposes restrictions in two ways:
 1. The variables that are important to predict the outcome.
 2. The functional form used to combine these variables to predict the outcome.
- **Reduce-form Approach:** Far less restricted.
 1. Relatively free choice of variables that might be useful (up to the modeller) in predicting outcomes.
 2. Relatively free choice of functional form that might be useful in predicting outcomes.

Structural and Reduced Form Approaches

- Reduced form approaches generally have been better than structural approaches.
- Especially in the area of corporate default forecasting.
- Reduced-form models have a great deal more “flexibility” – both in terms of variable choice, and functional form – thus making it a great target for big data tools.
- This lecture:
 1. Discuss and set-up a basic structural model of corporate default.
 2. Compare to the performance of reduced-form models.

Structural Models

- The most popular structural model of default is the [Merton distance to default](#) model.
- A straightforward application of option pricing.
- Core of the model: If a value of a firm's (or individual's) asset goes below a given level ("default barrier"), then the firm (individual) will not be able to pay their debts. → **Default!**
- In essence, these types of models have the same basic ingredients:
 1. Some stochastic process for the value of the firm's assets.
 2. Some default barrier which is the safe level of assets that guarantees repayment.
 3. A default time that arises when the assets fall below the default barrier.

The Merton Model

Assumptions:

1. The value of the firm's assets follows a Geometric Brownian Motion (GBM), under the physical measure:

$$\frac{dV}{V} = \mu dt + \sigma_v dW,$$

where:

- V is the total value of the firm,
 - μ is the expected instantaneous return on V ,
 - σ_v is the volatility of firm value, and
 - dW is a standard Wiener process.
2. The firm has a simple capital structure – issues equity, and one zero-coupon bond with face value F , which matures at some time T .

The Merton Model: Distance to Default

- At time T , either:
 - the firm has an asset value V that exceeds F , in which case the firm is solvent, the creditors are paid, and the shareholders take the residual value.
 - the firm has an asset value V that is less than F , in which case the firm is insolvent, defaults on the obligation to the creditors, and the shareholders get nothing.
- Shares/equity are a call option on the firm's assets, with a strike price of F .
- Now, **default** is the event that at time T , $V < F$.
- Aim is to find:

$$\Pr[V_T < F] = \Pr[\ln V_T < \ln F].$$

The Merton Model: Distance to Default

- Assuming V follows a geometric brownian motion, helps reformulate the probability of default (Vassalou and Xing, 2004):

$$\Pr\left(\ln\frac{V}{F} < 0\right) = \Phi\left(-\frac{\ln\frac{V}{F} + \left(\mu - \frac{\sigma_v^2}{2}\right)T}{\sigma_v\sqrt{T}}\right)$$

- Where $\Phi(\cdot)$ is the cumulative standard normal distribution function.
- In words: Probability of default depends on the firm's leverage.
- If leverage is high relative to its standard deviation, then the firm is expected to default.
- Challenge: we don't know V or σ_v since they aren't observable (need to make assumptions about μ). How do we find them?

The Merton Model: Equity Value of a Firm

- Equity value = Call option on the value of the firm.
 - Strike price of the call option is the face value of the firm's debt; maturity of the option is the maturity of the debt, T .
 - So, mathematically:

$$E = V\Phi(d_1) - e^{-rT}F\Phi(d_2),$$

where:

- E is the market value of the firm's equity,
- F is the face value of the firm's debt,
- r is the instantaneous risk-free rate,
- as before, $\Phi(\cdot)$ is the cumulative standard normal distribution function,
- $d_1 = \frac{\ln(\frac{V}{F}) + (r + \frac{\sigma_v^2}{2})T}{\sigma_v \sqrt{T}}, d_2 = d_1 - \sigma_v \sqrt{T}.$

The Merton Model

- And finally, one important equation relates the volatility of equity to the volatility of firm value:

$$\sigma_E = \frac{V}{E} \Phi(d_1) \sigma_v.$$

- The approach that is most commonly used is to take the two equations (for E and σ_E), and solve for the two unknowns (V and σ_v) by iterating to match observable values; then plug these values in to the formula for $\Pr(\ln \frac{V}{F} < 0)$ to get the probability of default.
- Bharat and Shumway (2008) also try just approximating these quantities without iterating, with good results.

The KMV Model

- Widely used approach in practice: KMV corporation (now sold to Moody's), computes proprietary Expected Default Frequencies (EDF) using this approach.
- The KMV model mainly differs in its final step from the method described here, i.e., mapping EDF to the distance to default measure, which may not come from a normal distribution...
- KMV uses a proprietary (expensive) historical database to map realized default frequencies to the distance to default.
- How good are these methods in practice?

Reduced-form Approach to Credit Risk

Reduced-form Approaches

- Reduced form approaches are at the opposite end of the spectrum.
- One approach is in Campbell, Hilscher, and Szilagyi (2008), and Bharat and Shumway (2008).
- Authors simply run a logit model on a set of observable firm characteristics.

Table III
Logit Regressions of Bankruptcy/Failure Indicator
on Predictor Variables

This table reports results from logit regressions of the bankruptcy and failure indicators on predictor variables. The data are constructed such that all of the predictor variables are observable at the beginning of the month over which bankruptcy or failure is measured. The absolute value of z-statistics is reported in parentheses. *denotes significant at 5%, **denotes significant at 1%.

Dependent variable: Sample period:	Model 1			Model 2		
	Bankruptcy 1963–1998	Failure 1963–1998	Failure 1963–2003	Bankruptcy 1963–1998	Failure 1963–1998	Failure 1963–2003
<i>NITA</i>	−14.05 (16.03)**	−13.79 (17.06)**	−12.78 (21.26)**			
<i>NIMTAAVG</i>				−32.52 (17.65)**	−32.46 (19.01)**	−29.67 (23.37)**
<i>TLTA</i>	5.38 (25.91)**	4.62 (26.28)**	3.74 (32.32)**			
<i>TLMTA</i>				4.32 (22.82)**	3.87 (23.39)**	3.36 (27.80)**
<i>EXRET</i>	−3.30 (12.12)**	−2.90 (11.81)**	−2.32 (13.57)**			
<i>EXRETAVG</i>				−9.51 (12.05)**	−8.82 (12.08)**	−7.35 (14.03)**
<i>SIGMA</i>	2.15 (16.40)**	2.28 (18.34)**	2.76 (26.63)**	0.920 (6.66)**	1.15 (8.79)**	1.48 (13.54)**
<i>RSIZE</i>	−0.188 (5.56)**	−0.253 (7.60)**	−0.374 (13.26)**	0.246 (6.18)**	0.169 (4.32)**	0.082 (2.62)**
<i>CASHMTA</i>				−4.89 (7.96)**	−3.22 (6.59)**	−2.40 (8.64)**
<i>MB</i>				0.099 (6.72)**	0.095 (6.76)**	0.054 (4.87)**
<i>PRICE</i>				−0.882 (10.39)**	−0.807 (10.09)**	−0.937 (14.77)**
Constant	−15.21 (39.45)**	−15.41 (40.87)**	−16.58 (50.92)**	−7.65 (13.66)**	−8.45 (15.63)**	−9.08 (20.84)**
Observations	1,282,853	1,302,564	1,695,036	1,282,853	1,302,564	1,695,036
Failures	797	911	1,614	797	911	1,614
Pseudo- R^2	0.260	0.258	0.270	0.299	0.296	0.312

Reduced-form Approaches

- These authors also add the Merton distance to default indicator into the set of variables, as do Bharat and Shumway (2008).
- Bharat and Shumway use a Cox proportional hazard model instead of a logit model (more on this in the second half of this course).
- Both authors find that the reduced form model significantly outperforms the structural model at all horizons.
- You might not find this very surprising, especially because of our previous lecture about model flexibility.
- But, the structural model performs very well, just not better than alternate approaches.

Campbell, Hilscher, Szilagyi (2008):

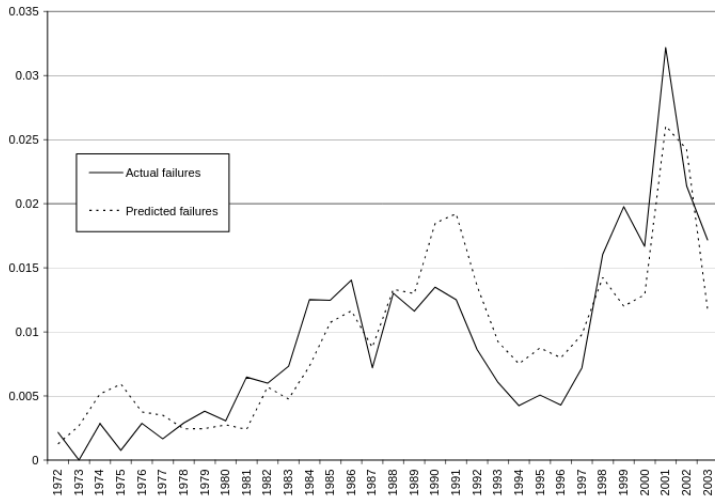
Table V
Distance to Default and Our Best Model

We report the coefficients on the “distance to default” (DD) variable in a logit regression by itself as well as when included in our best model (model 2 in Table III). The dependent variable is failure and the sample period is 1963 to 2003. Regression results are reported for various horizons: 0, 12, and 36 months. Panel A reports regression coefficients and the corresponding z-statistics (in parentheses). *denotes significant at 5%, **denotes significant at 1%. Panel B reports the in-sample and out-of-sample pseudo- R^2 statistics for the regressions from Panel A.

Lag (Months)	0	12	36
Panel A. Coefficients			
<i>DD</i> only	-0.883 (39.73)**	-0.345 (33.73)**	-0.165 (20.88)**
<i>DD</i> in best model	0.048 (2.62)**	-0.0910 (7.52)**	-0.090 (8.09)**
Observations	1,695,036	1,565,634	1,208,610
Failures	1,614	1,968	1,467
Panel B. R^2			
In-sample (1963 to 2003)			
<i>DD</i> only	0.159	0.066	0.026
Best model	0.312	0.114	0.044
<i>DD</i> in Best model	0.312	0.117	0.045
Out-of-sample (1981 to 2003)			
<i>DD</i> only	0.156	0.064	0.025
Best model	0.310	0.108	0.039

Structural and Reduced Form

Campbell, Hilscher, Szilagyi (2008):



Corporate vs Retail Default

- The reduced form approach looks like it achieves better predictive outcomes than the structural approach for corporate debt.
- What about retail default risk? Plenty of work in this area in practice (and less in academia).
- Big credit rating agencies (Equifax, Experian, CallCredit, ClearScore) use large amounts of data, including on past payment behaviour and demographics to generate a credit score for each individual.
- Tutorial exercises (and the guest lecture, TBC) will be on using publicly available retail finance datasets to predict retail default.

Applying ML Tools for Default Prediction

Set-up some more basic ML concepts

- Formalising Type I and Type II Errors.
- ROC Curves.
- SMOTE, Undersampling and Oversampling.
- (Next Lecture) Variable Selection and the LASSO.

Type I and Type II Errors

- Remember we are thinking through classification problems, in which we are classifying firms/individuals as likely to default or not to default.
- One good way to evaluate our prediction accuracy is to create a [confusion matrix](#), as we did while discussing LDA last week:

Predicted Class	True Outcome : Customers Default or Not	
	Positive (or Good)	Negative (Bad)
Positive (or Good)	True Positives	False Positives (Type I Error)
Negative (Bad)	False Negatives (Type II Error)	True Negatives

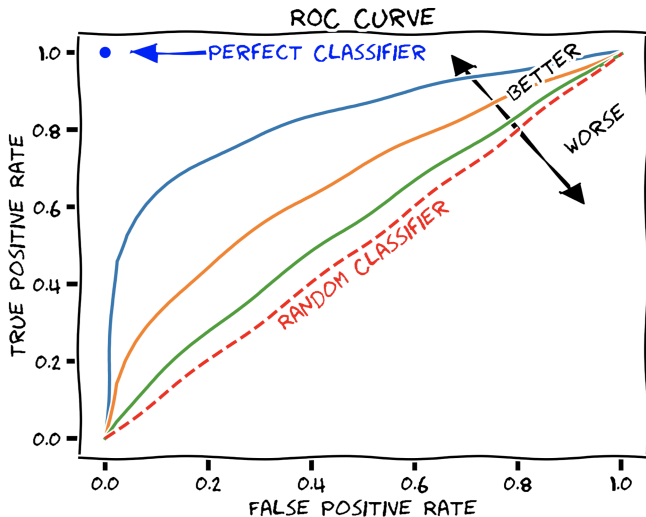
Type I and Type II Errors

- Clearly, in the default problem, false positives (Type I errors) are more costly than false negatives (Type II errors).
- We could set the threshold for classification as a defaulter lower, thus being sure to reduce Type I errors (We did this in the last lecture!).
- The cost of this, of course, is that Type II errors will inevitably increase.
- We can set the probability cutoff for classifying potential defaulters as high or low as we wish.
- Each choice will generate a different confusion matrix. How then, to choose?

ROC Curve

- Visualizes how changing the classification threshold affects Type I and Type II errors.
 - Roots in WWII.
 - “Receiver Operating Characteristic” ROC curves developed by the British as part of the “Chain Home” radar system.
 - ROC analysis was used to analyze radar data to differentiate between enemy aircraft and signal noise (e.g. flocks of geese)!
- What does it do?
 - Plots how true positive rate (% of defaulters correctly classified) varies with false positive rate (% of non-defaulters incorrectly classified) as threshold varies (not plotted).
 - Area under curve (AUC) is criterion to be maximized.

ROC Curve



Undersampling and Oversampling

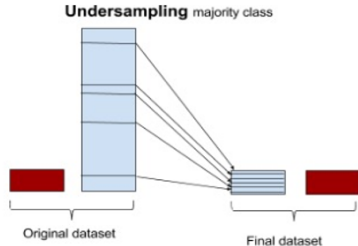
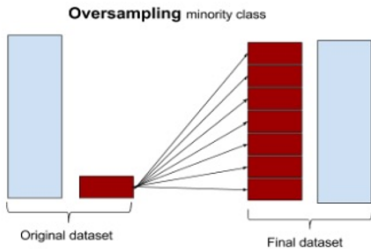
- On important issue in default classification problems is that there are very few actual defaults.
- Means that a simple approach in which we classify **all** observations as non-defaulters has very high accuracy!
- Clearly, we can use the ROC curve and varying thresholds in classifiers to do better than this **null** classifier.
- But this does suggest a deeper issue. One strategy to combat this is to judiciously use **undersampling** and **oversampling**.

Undersampling and Oversampling

- One possibility is to **undersample** the majority class.
- So if there are, say, 1,000 default observations and 100,000 non-default observations, we can randomly sample (with replacement) 1,000 observations from the distribution of non-default observations, and then train the algorithm on these data.
- We can also **oversample** the minority class, i.e., draw with replacement from the population of defaulters, until it becomes close to the size of the population of non-defaulters, and train the algorithm on this dataset.
- Variations include sampling closer to, or further away from, the decision boundary.

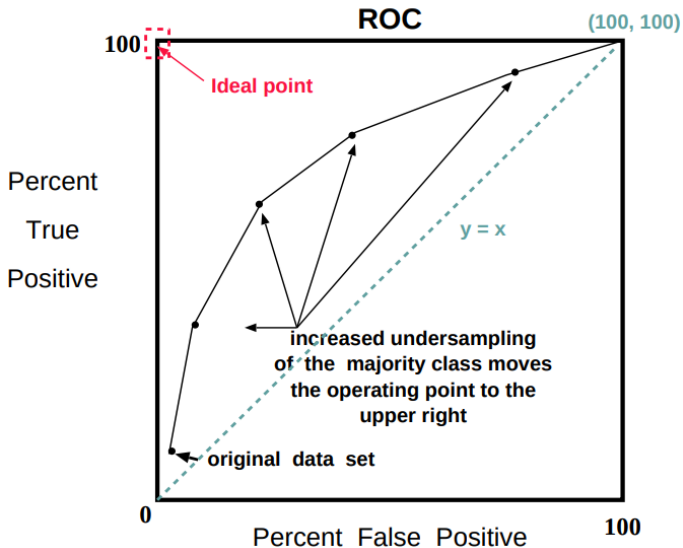
- SMOTE = Synthetic Minority Oversampling Technique.
- A similar approach (see Chawla et al. 2002), in which you oversample from the minority group by creating "synthetic" defaults.
- Create the synthetic default by:
 1. Randomly selecting a minority observation.
 2. Pick its nearest neighbor in feature space (right-hand-side variables).
 3. Compute the difference between the two vectors of features.
 4. Multiply this difference by a random number between 0 and 1, and add it to the randomly selected minority observation.
 5. This is the synthetic default - assign it the newly generated feature vector and a default indicator.
 6. Proceed to train the algorithm as usual.
 7. Can combine with undersampling the majority class.

Under and Over Sampling



Caution: This technique is only applied to the training data!

SMOTE and the ROC Curve



Barboza, Kimura and Altman (2017)

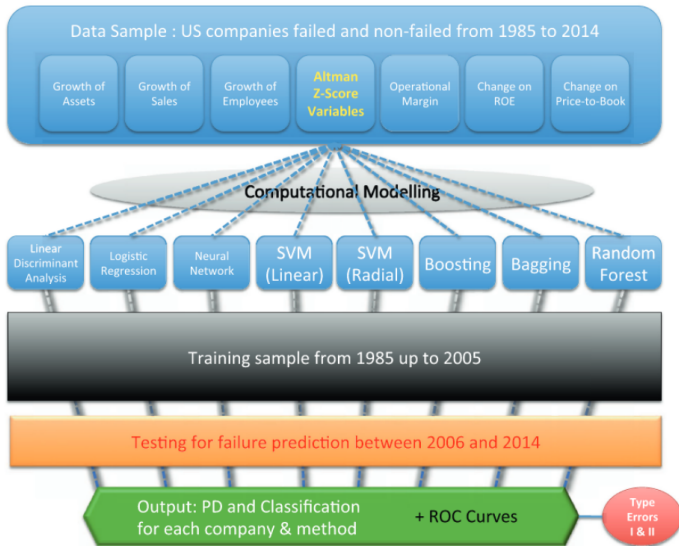
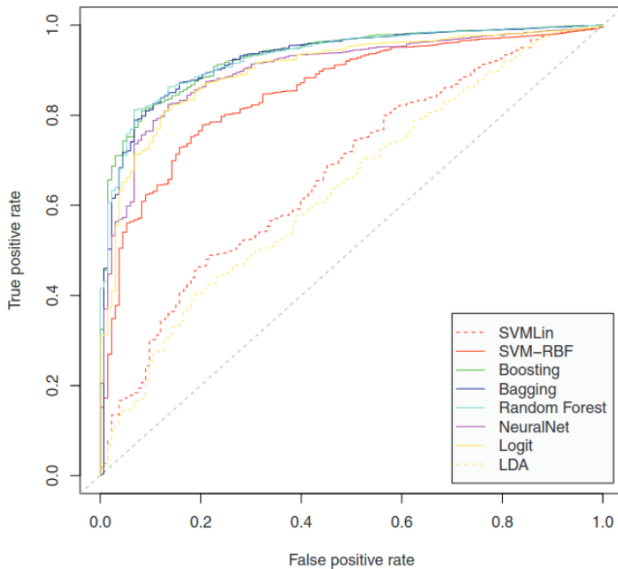


Table 3

Predictive variables of the default classification model for one year prior to bankruptcy. If bankruptcy filing occurred less than one semester after fiscal year end, data were collected from the previous fiscal year. Some measures require changes over time to compute these variables; we use data up to three years prior to bankruptcy.

Variable	Formula
X1	$\frac{\text{Net working capital}}{\text{Total assets}}$
X2	$\frac{\text{Retained earnings}}{\text{Total assets}}$
X3	$\frac{\text{Earnings before interest and taxes}}{\text{Total assets}}$
X4	$\frac{\text{Market value of share} * \text{number of shares}}{\text{Total debt}}$
X5	$\frac{\text{Sales}}{\text{Total assets}}$
OM	$\frac{\text{Earnings before interest and taxes}}{\text{Sales}}$
GA	$\frac{\text{Total assets}_t - \text{Total assets}_{t-1}}{\text{Total assets}_{t-1}}$
GS	$\frac{\text{Sales}_t - \text{Sales}_{t-1}}{\text{Sales}_{t-1}}$
GE	$\frac{\text{Number of employees}_t - \text{Number of employees}_{t-1}}{\text{Number of employees}_{t-1}}$
CROE	$\text{ROE}_t - \text{ROE}_{t-1} \text{ where } \text{ROE} = \frac{\text{Net income}}{\text{Common Stockholders' equity}}$
CPB	$\text{Price-to-Book}_t - \text{Price-to-Book}_{t-1} \text{ where } P/B = \frac{\text{Market value per share}}{\text{Book value per share}}$



Wrap-up

- We have seen a summary of structural and reduced form approaches to corporate default prediction.
- Bottom line seems to be that reduced form models do better at default forecasting for corporates.
- Retail default prediction has always been very reduced form.
- Also seen a few techniques that we might use to understand how best to apply machine learning to default prediction.
- Guest lecture and assignment an attempt to apply (and get you to apply) these insights on real-world datasets.

References

1. Barboza, Flavio, Herbert Kimura, and Edward Altman. "Machine learning models and bankruptcy prediction." *Expert Systems with Applications* 83 (2017): 405-417.
2. Bharath, Sreedhar T., and Tyler Shumway. "Forecasting default with the Merton distance to default model." *The Review of Financial Studies* 21.3 (2008): 1339-1369.
3. Campbell, John Y., Jens Hilscher, and Jan Szilagyi. "In search of distress risk." *The Journal of Finance* 63.6 (2008): 2899-2939.
4. Chawla, Nitesh V., et al. "SMOTE: synthetic minority over-sampling technique." *Journal of artificial intelligence research* 16 (2002): 321-357.
5. Vassalou, M., and Y. Xing. "Default risk in equity returns." *Journal of Finance*. (2004).