Big Data in Finance

Credit

Tarun Ramadorai

Default

- An important application of machine learning approaches is in the forecasting of (corporate and retail) default.

Default

- An important application of machine learning approaches is in the forecasting of (corporate and retail) default.

- Predicting default is important for many reasons.

Default

- An important application of machine learning approaches is in the forecasting of (corporate and retail) default.

- Predicting default is important for many reasons.
  - Understanding the risk of individual loans to optimally build a loan portfolio.

Default

- An important application of machine learning approaches is in the forecasting of (corporate and retail) default.

- Predicting default is important for many reasons.
  - Understanding the risk of individual loans to optimally build a loan portfolio.
  - Pricing these loans, since risk should equal expected return.

Default

- An important application of machine learning approaches is in the forecasting of (corporate and retail) default.

- Predicting default is important for many reasons.
  - Understanding the risk of individual loans to optimally build a loan portfolio.
  - Pricing these loans, since risk should equal expected return.
  - Computing value at risk for a loan portfolio to satisfy regulatory requirements.

Default

- An important application of machine learning approaches is in the forecasting of (corporate and retail) default.

- Predicting default is important for many reasons.
  - Understanding the risk of individual loans to optimally build a loan portfolio.
  - Pricing these loans, since risk should equal expected return.
  - Computing value at risk for a loan portfolio to satisfy regulatory requirements.
  - Forecasting equity returns, since distress risk should also be reflected in stock returns.

Default

- An important application of machine learning approaches is in the forecasting of (corporate and retail) default.

- Predicting default is important for many reasons.
  - Understanding the risk of individual loans to optimally build a loan portfolio.
  - Pricing these loans, since risk should equal expected return.
  - Computing value at risk for a loan portfolio to satisfy regulatory requirements.
  - Forecasting equity returns, since distress risk should also be reflected in stock returns.
  - Predicting bond and CDS prices,

Default

- An important application of machine learning approaches is in the forecasting of (corporate and retail) default.

- Predicting default is important for many reasons.
  - Understanding the risk of individual loans to optimally build a loan portfolio.
  - Pricing these loans, since risk should equal expected return.
  - Computing value at risk for a loan portfolio to satisfy regulatory requirements.
  - Forecasting equity returns, since distress risk should also be reflected in stock returns.
  - Predicting bond and CDS prices,
  - and much more....!

Structural and Reduced Form Approaches

- Two main approaches used to predict (corporate) default (with analogies in individual default forecasting):
  - Structural approaches to default forecasting.
  - Reduced-form approaches to default forecasting.

- When we call something a structural approach, this means that there is an underlying economic model, which imposes relatively tight restrictions on two things.
  - The specific variables that are important to predict the outcome.
  - The specific functional form used to combine these variables to predict the outcome.

- In contrast, a reduced form approach is far less restricted, so:
  - Relatively free choice of variables that might be useful (up to the modeller) in predicting outcomes.
  - Relatively free choice of functional form that might be useful in predicting outcomes.

Structural and Reduced Form Approaches

- As we will see, reduced form approaches have generally been better than structural approaches at predicting outcomes, especially in the area of corporate default forecasting.

- And as you have probably guessed, the flexibility of reduced form models of default makes them a great target for machine learning models.

- In this lecture, we will discuss and set up a basic structural model that has been quite popular for predicting corporate default, and then see how this compares to the performance of reduced-form models.

- Your assignment takes a reduced-form approach, and asks you to explore how well machine learning models do at this task.

Structural Models

- The most popular structural model of default is the Merton distance to default model. The model is a straightforward application of option pricing.

- All structural models of default essentially have the same central idea. If the value of the firm's (or individual's) assets go below a given level (the default barrier, the firm (or individual) will not be able to pay their debts, and will therefore default.

- In essence, these models have the same basic ingredients:
  - Some stochastic process for the value of the firm's assets.
  - Some default barrier which is the safe level of assets that guarantees repayment (default barrier).
  - A default time that arises when the assets fall below the default barrier.

The Merton Model

- The most popular structural model of default is the Merton distance to default model. The model is a straightforward application of option pricing.

- The model makes two important assumptions. First, 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.

- Second, 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.

- Another way to put it is that 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, $V < F$. So we wish to find:

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

The Merton Model: Distance to Default

- But given the assumption of V following a GBM, it turns out that this is (see Vassalou and Xing, 2004, JF for details):

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

  - Where $\mathrm{N}(.)$ is the cumulative standard normal distribution function.

- Intuition: notice that the probability of default depends on the firm's leverage If leverage is high relative to its standard deviation, 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

- In the Merton model, the equity value of the firm is a 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:

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

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, $N(.)$ 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:

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

- 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.

- Key question: How good is the method in practice?

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%.

	Model 1			Model 2		
Dependent variable:	Bankruptcy	Failure	Failure	Bankruptcy	Failure	Failure
Sample period:	1963–1998	1963–1998	1963–2003	1963–1998	1963–1998	1963–2003
NITA	−14.05	−13.79	−12.78			
	(16.03)**	(17.06)**	(21.26)**			
NIMTAAVG				−32.52	−32.46	−29.67
				(17.65)**	(19.01)**	(23.37)**
TLTA	5.38	4.62	3.74			
	(25.91)**	(26.28)**	(32.32)**			
TLMTA				4.32	3.87	3.36
				(22.82)**	(23.39)**	(27.80)**
EXRET	−3.30	−2.90	−2.32			
	(12.12)**	(11.81)**	(13.57)**			
EXRETAVG				−9.51	−8.82	−7.35
				(12.05)**	(12.08)**	(14.03)**
SIGMA	2.15	2.28	2.76	0.920	1.15	1.48
	(16.40)**	(18.34)**	(26.63)**	(6.66)**	(8.79)**	(13.54)**
RSIZE	−0.188	−0.253	−0.374	0.246	0.169	0.082
	(5.56)**	(7.60)**	(13.26)**	(6.18)**	(4.32)**	(2.62)**
CASHMTA				−4.89	−3.22	−2.40
				(7.96)**	(6.59)**	(8.64)**
MB				0.099	0.095	0.054
				(6.72)**	(6.76)**	(4.87)**
PRICE				−0.882	−0.807	−0.937
				(10.39)**	(10.09)**	(14.77)**
Constant	−15.21	−15.41	−16.58	−7.65	−8.45	−9.08
	(39.45)**	(40.87)**	(50.92)**	(13.66)**	(15.63)**	(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 later in the course).

- Both authors find that the reduced form model significantly outperforms the structural model at all horizons.

- You might not find this very surprising! However, notably, the structural model performs very well.

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 regregression 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			
DD only	−0.883	−0.345	−0.165
	(39.73)**	(33.73)**	(20.88)**
DD in best model	0.048	−0.0910	−0.090
	(2.62)**	(7.52)**	(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)			
DD only	0.159	0.066	0.026
Best model	0.312	0.114	0.044
DD in Best model	0.312	0.117	0.045
Out-of-sample (1981 to 2003)			
DD only	0.156	0.064	0.025
Best model	0.310	0.108	0.039

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.
- The use a reduced form modelling approach. Fast-growing. Also uses clicking behaviour etc. on website shopping to offer you "instant credit".
- Exercises (and the guest lecture, soon) will be on using publicly available retail finance datasets to predict retail default.

Applying Machine Learning to Default Prediction

- This whole area seems like an obvious place where ML might result in big improvements.
- Need to think through a few issues a bit more carefully though.
- ROC Curves and Understanding Type I and Type II Errors.
- 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 or individuals as likely to default or not to default.

- One good way to evaluate our prediction accuracy is to create a confusion matrix:

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.

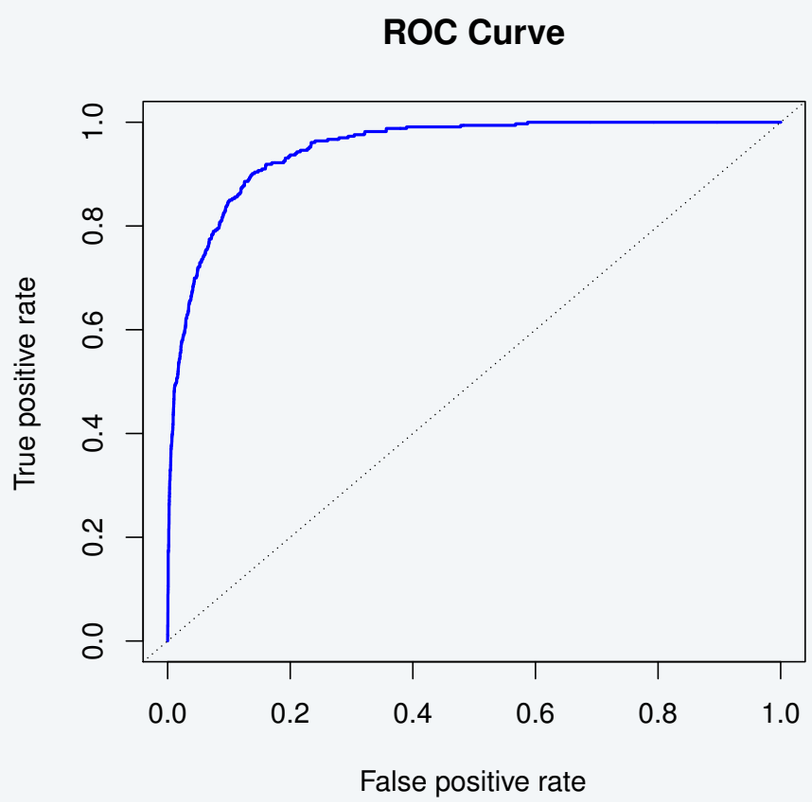
- The cost of this, of course, is that Type II errors will inevitably increase.

- For concreteness, think of a case in which we are using a logistic regression classifier.

- 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.

ROC Curve

- Visualizes how changing the classification threshold affects Type I and Type II errors. ("Receiver Operating Characteristic" - use in radio signal detection.)
  - 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.



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 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

- Stands for 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:
  - Randomly selecting a minority observation.
  - Pick its nearest neighbor in feature space (right-hand-side variables).
  - Compute the difference between the two vectors of features.
  - Multiply this difference by a random number between 0 and 1, and add it to the randomly selected minority observation.
  - This is the synthetic default - assign it the newly generated feature vector and a default indicator.
  - Proceed to train the algorithm as usual.
  - Can combine with undersampling the majority class.

Conclusion

- 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 (now February 3) and assignment an attempt to apply (and get you to apply) these insights on real-world datasets.

Imperial College
Business School

Imperial means
Intelligent Business