February 2021

Modelling the Default of Private Student Loans

This article presents a survey of methods for modelling student loan default.

Modelling the Default of Private Student Loans

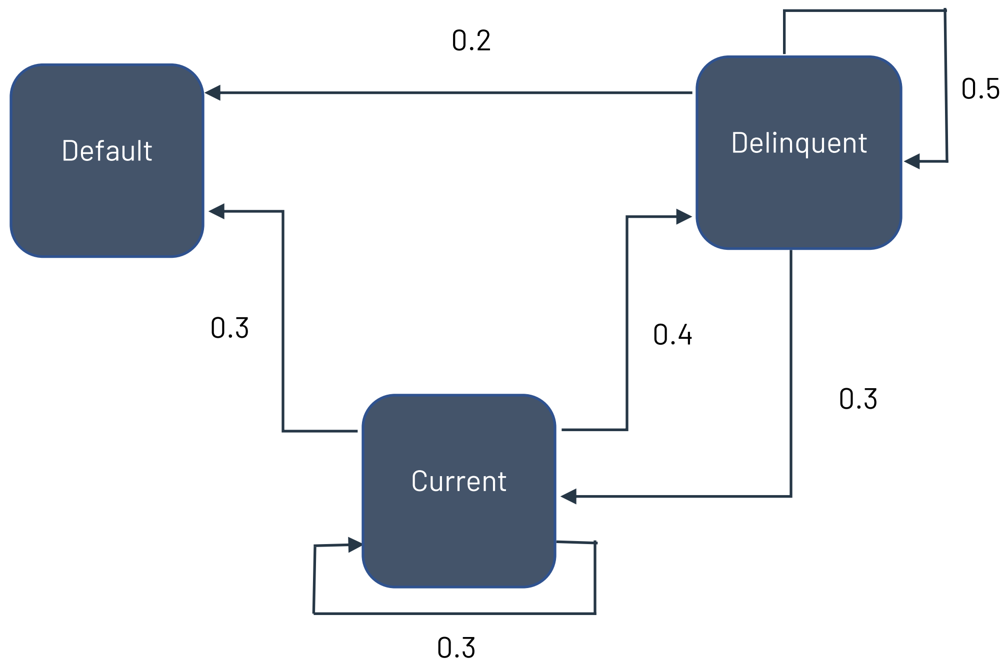
Given the rising prominence of private student loans, formulating and estimating a model of student loan default remains an important goal. This article summarizes several approaches to modeling student loan default. While not an exhaustive list, the options below are a useful summary of many standard approaches in predicting loan default of private student loans. These options differ in complexity and ease of implementation, and the choice of which approach is the best to use ultimately depends on the requirements of the case or consulting engagement, and the available data.

Roll Rate Analysis

A roll rate model is a simple example of a Markov Chain model, which describes how a random variable can transition from one state to another. In the context of student loans, the status of the loan (current, delinquent, in default, etc.) is the random variable that transitions between different states, or the different possible statuses that a student loan can be in. Thus, the main aim of the model is to calculate the set of probabilities with which a loan will transition from any given status to another.

This set of probabilities transitioning from one state to another are known as the *transition probabilities* and is represented by a matrix, known as a *transition matrix.* Equipped with a transition matrix, the status of a loan, including default, could be predicted in the next period. Take for instance the example model depicted in **Figure 1:** *Illustration of Roll Rate Analysis*. A loan can be in any of the three states—default, 30-day delinquent, or current—and the number above each arrow indicates the probability that the loan transitions from one state to another. For instance, for a loan that is current this period, the probability that it remains current next period is 30 percent, the probability that will become 30-day delinquent next period is 40 percent, and the probability that it will default next period is 30 percent.

Figure 1: *Illustration of Roll Rate Analysis*



To calculate the set of transition probabilities in a roll rate model, a researcher must use data points over at least two periods of time in order to measure the rate of transitions between the states. Other than that, data needs are minimal, as this requires only which state a loan is in in for each of the two periods. Equipped with this information, the set of probabilities required for roll rate analysis can be estimated. Further, this model requires very few modelling assumptions, as it represents a reduced-form empirical estimate of the actual transitions that occur for loans.

Due to the straightforward nature of this model, and its minimal requirements in terms of assumptions and data, a roll rate analysis can be an easy way to model student loan default.

Regression Models

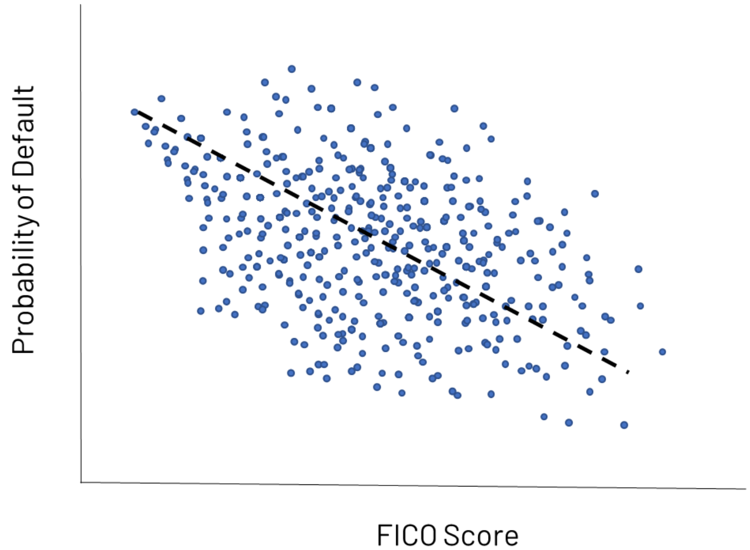
Regression analysis is a powerful and widely used statistical tool. Regressions consist of two main ingredients: a response variable (also called dependent variable) and a set of explanatory variables (also called independent variables). Regression studies the statistical relationship between the response variable and the explanatory variables using data.

Here, the response variable is default of student loans, and explanatory variables might include:

1) *Individual-level information* such as borrower's FICO score, amount of other debts, type of institution and program enrolled in, and expected employment and income;
2) *Loan-level information* such as loan amount and terms; and
3) *Aggregate macroeconomic variables* such as cost of living, expected economic growth, and expected interest rates.

Having learned about the relationship between the explanatory variables and the response variable, the researcher can predict what values the response variable would take, given particular values of the explanatory variables. For example, if we know a borrower's credit score, loan amount, income, and other relevant information, we can use a regression model to predict how likely it is that the borrower will default. **Figure 2**: *Illustration of Regression Methodology* shows how an illustration of a regression could relate a borrower's FICO Score to the probability of default.

Figure 2: *Illustration of Regression Methodology*



Regression analysis generally requires cross-sectional data with the variables that the researcher deems to be important in predicting the response variables. In some cases, when it is not possible to observe certain variables, the researcher can use *proxy variables*, which may not by themselves be relevant but which serve in place of the unobservable variable that they shadow. For instance, we don't have a way to directly measure a particular borrower's credit-risk, but the borrower's FICO Score generally provides a good indicator of the borrower's underlying credit-risk.

Regression analysis also requires some assumptions on the part of the researcher. Initially, the researcher has to make assumptions regarding the mathematical relationship between the explanatory variables and the response variables. These assumptions determine the exact specification, or flavor, of the model that is used. Furthermore, the researcher also has to make assumptions regarding the causal chain between the variables included in the model.

If the data and assumption requirements outlined above are met, a regression can be a very useful tool for predictive analysis. Using alternative values of the explanatory variables, the researcher can measure the expected change in the response variable when one, or all the explanatory variables are changed.

Regression can also be useful beyond simple prediction. It is a particularly powerful tool for making inferences regarding the factors that influence the response variable. When explanatory variables are included in the model, the model calculates a *coefficient*, which tells the researcher the direction, magnitude, and significance of the impact of that explanatory variable on the response variable. Significance, in the statistical context, refers to a statistical relationship that is so strong that it is unlikely to be due to chance alone. The significance of a variable may indicate the relevance of a particular factor. For instance, one could test the significance of various factors such as the co-signer's credit score, or the servicer, or macroeconomic variables, on the probability of default. Such tests could provide valuable analysis in the context of litigation, where one may be interested in knowing whether or not a particular factor had a causal impact on default.

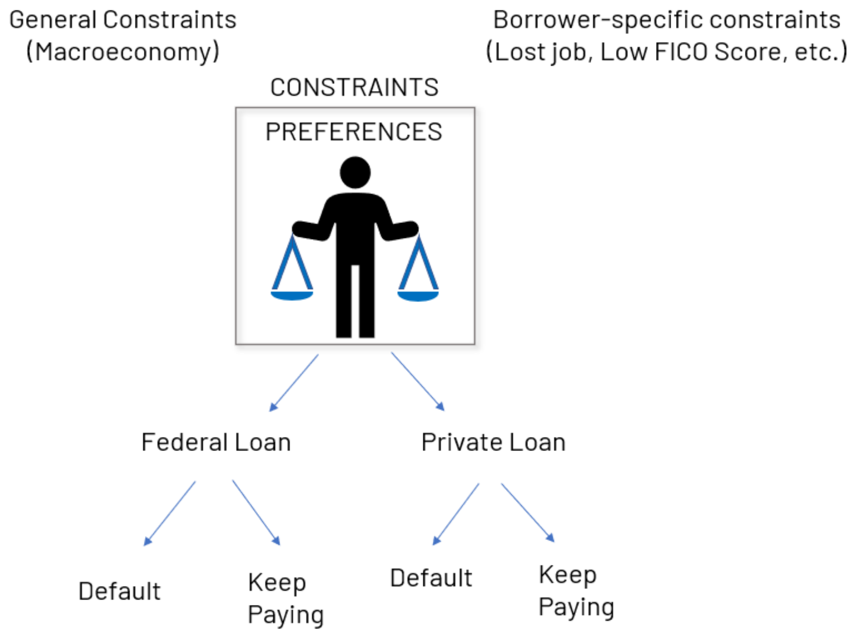
However, caution should be exercised in blindly treating significance as an indication of relevance, and particularly in attributing causality. While regression models are a potent tool, they should be used with ample common sense, and with an understanding of the assumptions underlying them. Here again, the particular question plays a crucial role in determining whether regression is the appropriate modelling choice, and the extent to which regression can be leveraged to provide critical insights.

Structural Models

Structural models are like mini-stories in which individuals (called "economic agents") live a hypothetical life and make hypothetical decisions. These mini-stories are constructed to mimic the real world and the incentives and constraints individuals face when they make decisions. Structural models state clearly what individuals are trying to achieve, what resources they have at their disposal, how they interact with each other, and ultimately what decisions they make given these goals and constraints. In other words, the models lay out a set of rules by which different variables interact, and the modeler follows the rules to their logical conclusion.

Structural models generate predictions through counterfactual analysis. Given the rules of the model, the researcher can change information relevant to individuals' decision making and observe how individuals react and how these reactions affect predictions from the model. For instance, in a structural model that links interest rates and default probabilities, one could plug in alternative interest rates and see the expected changer in default rates.

Figure 3: *Illustration of Decision Making in a Structural Model*



Structural models require some data to calibrate the model, *i.e.*, to adjust the model such that it provides a sufficiently accurate mathematical relationship between actually observed data points. However, these data needs depend on the particular model, and on how much the researcher is willing to assume rather than calibrate with real data. Compared to both roll rate and regression models, the list of assumptions tends to be more elaborate for structural models, though these assumptions are generally stated explicitly such that they can be verified or rejected.

Like regression analysis, structural models can provide rich predictions and provide quantitative evidence about the relevance of various factors in predicting the response variable of interest. Additionally, structural models lay bare particular decision-making processes, which may also be valuable depending on the context. For instance, in the case of student loans, a researcher might model how the borrower chooses between the large menu of different student loans available, and then how the probability of default follows from this initial choice. Depending on the context, structural models provide a powerful, but complex, tool to answer key questions that may arise in the course of litigation or consulting.