Coursework 2- Credit Risk Analytics

Lending Club (https://www.lendingclub.com) is a well-known peer-to-peer (P2P) lenders operating in North America. Its business model is to let potential investors diversify their risk by splitting their investments across multiple loans. To add transparency to their models, Lending Club has been realising their lending history in its entirety, available for the period 2007 to 2018 for the United States, along with loan performance for them. The dataset has approximately 2.3 million loans and around 110 variables.

In this coursework, you will develop a fully compliant advanced IRB model from the data they make available, from the raw data to the level 2 calibration, using what you have learned in the lectures. The objective of the coursework is to estimate the capital requirements for Lending Club were they to be regulated (they are currently not).

1. (30%) Prepare the dataset to make it ready for a credit scoring application model and for a Loss Given Default model and calculate the default variable and the workout LGD variable. Discuss all your decisions, particularly focus on which variables can be used to predict PD, which ones to predict LGD, which ones are variables used to construct your objective variables, and which variables cannot be used. Note the variable "grade" is not a predictive variable, for example, as it includes the business logic by Lending Club.
2. (35%) Construct a scorecard which can model the probability of default for the loans. As you do not have information regarding 12-month performance, use the status "Chargeoff" as the objective variable from variable "loan_status". Discuss your choice of variables, your decisions regarding Weight of Evidence and other transforms you choose to make, and your final performance. Discuss the variable importance. How many variables do you recommend using?
3. (35%) Construct two Loss Given Default models, using Random Forest and XGBoosting, **over the defaulted loans only**. Use cross-validation to determine your optimal parameters, if necessary, discuss the variable importance and the accuracy metrics you see relevant. Compare the performance of both models and discuss your findings. Discuss the variable importance of both models. Do they agree? Why? Apply this model to the non-defaulters and discuss the average estimated LGD values over these cases.
4. (Extra credit, 20%. Maximum score 100%) Using the monthly macroeconomic information you consider relevant (see for example https://stats.oecd.org/Index.aspx), calibrate a long-run PD and downturn LGD model for the loans granted regressing your monthly Lending Club's PDs (from your objective variable) per rating (from the variable 'grade') against the macroeconomic variables. Use the long-term forecasts you can find online from reputable sources (for example the OECD) for your long-term calibrated values. If you cannot find them, assume a value which makes sense to you and explain why. For the downturn values, select the worst month GDP-growth-wise and use those macroeconomic values.

Conditions of the coursework

Software: You must use Python to run the numerical calculations over your portfolio. A copy of your jupyter notebook must be attached to the coursework as an appendix in readable format, and a link

to the notebook must also be included. Instructions how to export to PDF can be found here: https://stackoverflow.com/questions/52588552/google-co-laboratory-notebook-pdf-download

Word Limit: 2000 words +/-10% either side of the word count is deemed to be acceptable. Any text that exceeds an additional 10% will not attract any marks. The relevant word count *includes* items such as cover page, executive summary, title page, table of contents, tables, figures, in-text citations and section headings, if used. The relevant word count *excludes* your list of references and any appendices at the end of your coursework submission.

You should always include the word count (from Microsoft Word, not Turnitin), at the end of your coursework submission, before your list of references.

Title/Cover Page: You must include a title/ cover page that includes: your Student ID, Course Code, Assignment Title, Word Count. This assignment will be marked anonymously, please ensure that your name does not appear on any part of your assignment.

Submission Deadline: November 14th, 23:59.

Turnitin Submission: The assignment MUST be submitted electronically via OWL. All required papers may be subject to submission for textual similarity review to the commercial plagiarism detection software under license to the University for the detection of plagiarism. All papers submitted for such checking will be included as source documents in the reference database for the purpose of detecting plagiarism of papers subsequently submitted to the system. Use of the service is subject to the licensing agreement, currently between The University of Western Ontario and Turnitin.com (http://www.turnitin.com).

Late Submission: Late submissions are possible up to a week after deadline. There is a 10% penalty per day of late submission subtracted directly from the final mark. Submissions after the 7 days are not accepted and will be considered a non-submission.