

Problems

1. Data-preprocessing

- A. Explain qualitatively (beyond what you find on Chat GPT) the role of borrower income on credit risk prediction in a maximum of 200 words.
- B. Collect time series from 2001 to 2015 of borrower incomes by state in the US from <https://www2.census.gov/programs-surveys/cps/tables/time-series/historical-income-households/h08.xlsx>. Describe the data using the describe method in pandas and plot the time series using matplotlib. Interpret all outputs economically. Interpolate missing years
- C. Calculate growth rate of income by state and time. Describe the heterogeneity (differences) in growth rates using economic interpretations. Merge the time series by state and time with the mortgage data assuming that time=1 corresponds to 2001:Q1 and time=60 corresponds to 2015:Q4. Make assumptions if there are data challenges.

2. PD modelling

1.

- A. Estimate a basic credit risk model for mortgage default probabilities (PD). You may choose a logit or a probit model. Include two standard explanatory variables which are FICO and LTV at origination. Do not include further variables. Compute the estimated PD for all mortgage loans and periods. Plot the average probability of default by time in a chart. Provide your code, output for the model and interpret the output.
- B. Estimate the PD model again by including explanatory variables in part (a), and the state-level income growth from Question 1 in one regression. Compute the estimated PD for all mortgage loans and periods. Plot the average probability of default by period in a chart. Provide your code, an output for the model, the plots and interpret the output economically.
- C. Compare the accuracy of the two models from sub-questions 2a and 2b. Present and explain your findings with regard to model accuracy.

3. LGD modelling

1.

- A. Run a linear regression model to predict LGD. Include the same set of explanatory variables used in the question 2B. Compute the estimated LGD for all mortgage loans and periods. Plot the average LGD by period. Provide your code, output, plot and interpret the output economically.
- B. Explain how income may be a potential driver of LGD? Suggest, next to income and any other features you have used in 3A, two additional features that you think can explain for mortgage LGD. Explain the rationale (relation to LGD)

4. Expected loss calculation

Compute the level of expected one-period loss for all mortgage loans and past periods (i.e., in sample for periods 1 to 60). PD should be inferred from question 2b and LGD should be inferred from question 3b. You may set the exposure at default to one dollar so that risk measures can be interpreted in percentages. Plot the average expected loss by time period in a charts. Provide your code, plots and analyse the output economically.

5. Climate change

Please read the attached papers on climate change: [Basel and climate change Download](#) [Basel and climate change](#) and [Basel and climate change 2 Download](#) [Basel and climate change 2](#).

1.

- A. How does climate change impact borrower income? Collect a time series from 2001 to 2015 of one climate feature. Build a linear regression model explaining the income growth rates from 1C (by year only) using the climate features making suitable assumptions
- B. Collect a prediction of future values for your climate feature from 5A from 2015 onward
- C. Consider the PD and LGD over the lifetime of a 30-year mortgage loan (starting from the end of the data in 2015 until 2045). Calculate the PD and LGD term structure (by time) for a mortgage loan of a representative borrower and loan feature predictions of 5B using the model that you have developed in 5A.