**Panel Data Modelling Documentation**

Notation: file name that are mentioned for the first time is highlighted in bold; file name that are subsequently mentioned is set to italics.

* Overview: the panel data modelling steps are carried out in Python using the script file **Panel data modelling.py** and data file **Final rating dataset.csv**.
* Setup: when loading the data from *Final rating dataset.csv*, the gvkey of the company and the year information is set as the index, which is for the use of panel data analysis.

The regression analysis of the panel data consists of two main parts: univariate and multivariate modelling. All the modelling steps use the column of credit rating as the dependent variable.

1. Univariate modelling: both Pooled OLS model and Fixed Effects model are used on all the macro predictors (economic predictors and weather-related predictors)
   1. In the case of Pooled OLS model, only 1 macro predictor is used as the independent variable;
   2. In the case of Fixed Effects model, apart from the 1 macro predictor used, four financial ratios (as mentioned in **Data Processing Documentation.docx** 3b) of the companies are also used as the control variables;
   3. After filtering and removing models with R-squared that are less than 0.058 (the R-squared range of the univariate result is between 1.95E-06 and 0.0599), and further filtering and removing models with p-values that are higher than 0.03, fourteen (14) variables remained (a detailed list of the 14 variables can be found on the presentation slides).
   4. These 14 variables are later further selected using multivariate analysis by choosing a model that achieve the highest adjusted R-squared value. It is also observed that Fixed Effects models generally outperform Pooled OLS models. Therefore, the multivariate analysis will focus on Fixed Effects models.
2. Multivariate modelling: all possible combinations with sizes ranging from 2 (macro predictors) to 5 (macro predictors) from the 14 variables are used in Fixed Effects modelling
   1. As the size of the combination increases, an increase in the adjusted R-squared of the models can be observed. However, the improvement of adding one more macro predictor is very little.
   2. The model with the highest adjusted R-squared value is built using the following 5 macro predictors: tropical\_storm\_death​, wildfire\_count, flood\_death, GDP\_per\_capita, and heat\_death (more detailed result of the best model can be found on the presentation slides or the output from the last “print” statement in the file *Panel data modelling.py*)
   3. The individual p-values of these macro predictors’ coefficients suggest that they are significant.