**Project #2 (AI for Finance)**

**Spring 2023**

**(Due 9 April, 2023)**

**Scenario: You are a portfolio manager. Hence, you want to rebalance your portfolio from high beta stocks to low beta stocks before recessions. As a result, you may want to find a model that has some predictability of recessions.**

1. **Try to read the paper by Estrella and Hardouvelis (1991) (**<http://hardouvelis.gr/wp-content/uploads/2017/12/JF_June_1991_462_555-76_HQ-1.pdf>**).**

**First, you construct the dataset.**

**The analysis is for the 1955-present sample.**

**Here are the data you need**

**1) the Treasury TERM spread;**

**2) NBER Recessions data.**

**The relevant data are available @ BEA/NBER – some data you already have/you know the source. Since GDP data is quarterly, and hence “recession” is quarterly data. We need to do the analysis quarterly. Convert monthly data to quarterly by taking an average.**

* **You need to run a Probit model as explained in class.**
* **Compare the results of the probit model with two AI models**
* **1) PNN (1960s): probabilistic neural network;**
* **2) MLP (1940s): multilayer perceptron.**

**Tips:**

* **In python there is a command in order to calculate lags.**
* **Plot recessions’s values and TERM value.**
* **For each auroc I will have 3 valued, one for probit, one for pnn and one for mlp**

**PART 1**

**The dependent variable is the NBER recessions binary variable. Since we want to forecast recessions, use lags (up to 12 quarters) of the explanatory variables. Report the results in a tabular form.**

**Use TERM as the explanatory variable – 12 models for 12 quarters lags of TERM, one lag at a time.**

* 1. **In each probit/logit model, you are going to see the model fitness in terms of R-sq. Report R-sq for the probit model.**
  2. **However, predictions are never “1” or “0”. Use a different evaluation technique: Area Under the ROC curve (AUROC). Random guess would have an AUROC value of 0.5. Any model with AUROC greater than 0.5 is better than a guess. Calculate AUROC of each model (12 AUROC values). Report those in a tabular form. Use the full sample. In econometrics, we call it in-sample analysis.**
  3. **Now use the AI and probit/logit prediction models. Compute AUROC values. Report those in a tabular form (you will have 36 AUROC values). Use 2/3 of data for training. Use 1/3 for evaluation. In econometrics, we call it pseudo-out-of-sample analysis.**
  4. **Compare the AUROC values and find out what model best predicts recession.**

**PART 2**

**Assume that if we are in a “Recession” today, the “Recession” today is a good predictor of recession tomorrow. That is, in each quarter, add a lag of “Recession” as a predictor. As a result, in each quarter, you have 2 explanatory variables – make sure both have the same lag, otherwise prediction is faulty.**

**Redo part 1, step c, with the new model.**

**The probit model**

Youestimate the probability of recessions using the following probit model as per the literature (e.g., Estrella and Hardouvelis 1991):

(1)

where V is a vector of predictor variables that includes TERM, etc. and *l* is the number of lags used and represents the forecast horizon. The dependent variable () is a binary variable, which is “1” if the economy is in a NBER recession quarter and ‘0’ otherwise. Models are evaluated based on Pseudo-R-Squared value and AUROC (mainly).

**In python, I have this datas inside two dataframe with periodindex of frq: Q-DEC**

**1) the Treasury TERM spread;**

**2) NBER Recessions data.**

**Now i need to run Probit Model, the dependent variable is the NBER recessions binary variable. Since I want to forecast recessions, use lags (up to 12 quarters) of the explanatory variables and Use TERM as the explanatory variable – 12 models for 12 quarters lags of TERM, one lag at a time. Do also the following:**

* + **In each probit/logit model, you are going to see the model fitness in terms of R-sq. Report R-sq for each probit model.**
  + **However, predictions are never “1” or “0”. Use a different evaluation technique: Area Under the ROC curve (AUROC). Calculate AUROC of each model (12 AUROC values). Report those in a tabular form. Use the full sample. In econometrics, we call it in-sample analysis.**
  + **Now use the AI and probit/logit prediction models. Compute AUROC values. Report those in a tabular form (you will have 36 AUROC values). Use 2/3 of data for training. Use 1/3 for evaluation. In econometrics, we call it pseudo-out-of-sample analysis.**

**Report the results in a tabular form**