SPARKASSE

MIM Econometrics
Homework 2: Non-Linear Models for Prediction
Prof. Jan Nimczik

You are a loan officer at the Sparkasse bank in Hamburg. Sitting on your desk are personal loan applications from 10 individuals (gender unknown): Maier, Ginter, Mauser, Beckmann, Heckmann, Buntschuh, Kremer, Schmidt, Oener, and Fassbender. When individuals apply for a personal loan, it is standard procedure to collect some basic information regarding their financial characteristics. The information available for these 10 loan applicants are available in the data set *prospects.csv*. In addition, you have at your disposal data on a random sample of 1000 exisiting bank clients, available in *credit.csv*. It includes data on the basic financial characteristics, as well as well as whether these clients defaulted on their loan (*default=1*) or not (default=0). Variable descriptions are provided in the Table below in this document. The datasets, *prospects.csv* and *credit.csv,* are available on the course website.

It is your job to decide whether or not grant or deny these loan applications. You are going to build an econometric model, estimate it using data on existing customers, and use these estimates to make predictions that will help you decide on whether or not to accept the new loan applications.

Please complete the following steps in your analysis:

1. Build and estimate a model to predict the probability that these applicants will default on their loan (if you choose to disburse one).
    i. Write down a sensible model; explain your economic intuition for including the variables on the right-hand side of this model.
    ii. Estimate this model using:
        a. Linear Probability Model
        b. Probit
        c. Logit

    Does the model make sense? (sign and significance of variables)
2. Using the estimates from part 1.ii (for the three models separately, and for each applicant separately), predict each of these applicants' probability of default.
3. Using the predicted default probabilities for each applicant you can now turn to making decisions on the applications. Each of these applications is for a 1-year loan, but for different amounts (see column "amount" in the prospects data). Assume that the annual interest rate that you can charge from your clients is 10%. Further assume that your cost of borrowing is zero and you have no other variable costs.
    - Whom would you lend to given the predicted default probabilities and the given interest rate? Show your calculations (and try to do them in R).
    - What is your expected profit if you follow this recommendation?

- Use the confidence interval in the linear probability model to calculate a negative case scenario: What is your expected profit if the default probabilities are at the upper value of the prediction interval

Variable Name	Variable Description
seniority	job seniority (years)
home	type of home ownership
age	client's age
marital	marital status
job	type of job
expenses	client's expenses
income	client's income
assets	client's assets
debt	client's existing debts
amount	loan amount (for prospects, this is the personal loan amount they have applied for)
savings	client's savings