

Statistical Analysis and Description

1 The general nature of the problem

The dataset that we are studying, is provided by the *Federal Reserve Economic Data*. It contains information of economical data in the USA from 01/04/1980 to 02/04/2000 on a weekly basis.

There are 1049 observations with 16 variables. This tell us that it is a multiple regression of supervised learning problem. The goal of the problem, is to deal with our dataset in order to find the best model fitting which is well-predict to our target variable "1-Month CD Rate".

However, we need firstly to check and clean up our dataset before dealing with the methodology to find the best fitting model. To facilitate in reading, we would like to label each variables as shown in the table below following by theirs description. Luckily, we know that there is no missing value or outliers detection in our data. Then, we are going to do the univariate and multivariate analysis of variables to see theirs properties and theirs relationship between each others.

Label	Description
1YM	1-year Treasury Constant Maturity rate
DDB	Demand Deposits at Commercial Banks
30YM	30-Year Conventional Mortgage Rate
EFF	Effective Federal Funds Rate
3MA	3-Month Treasury Bill Rate-Auction Average
M1S	M1 Money Stock
3MS	3-Month Treasury Bill Rate-Secondary Market
TCD	Total Checkable Deposits
3YM	3-Year Treasury Constant Maturity Rate
TLL	Total Loans and Leases of All Commercial Banks
5YM	5-Year Treasury Constant Maturity Rate
TSD	Total Savings Deposits
BCA	Bank Credit of All Commercial Banks
TWE	Trade Weighted Exchange Index Major Currencies
CCS	Currency Component of Money Stock
1MCD	1-Month CD Rate

2 The univariate analysis

The aim to do the univariate analysis, is the see the properties of each variables such as central tendency (mean, mode and median) and dispersion: range, variance, maximum, minimum, quartiles, and standard deviation. For example, the variable "1YM" has minimum about 77 and maximum around 143 while mean and median share quite similar value. Moreover, we observe that the mean and median of other variables have quite similar values too.

Since we have more than one variable in our data set. We would prefer to do also the multivariate analysis which work on correlated variables.

3 The multivariate analysis

As mention above that we have more than one variable in our data, thus it is really essential to plot the correlation matrix which can help us to study the relation between our target variables "1MCD" with other explanatory variables and either between couples of explanatory variables.

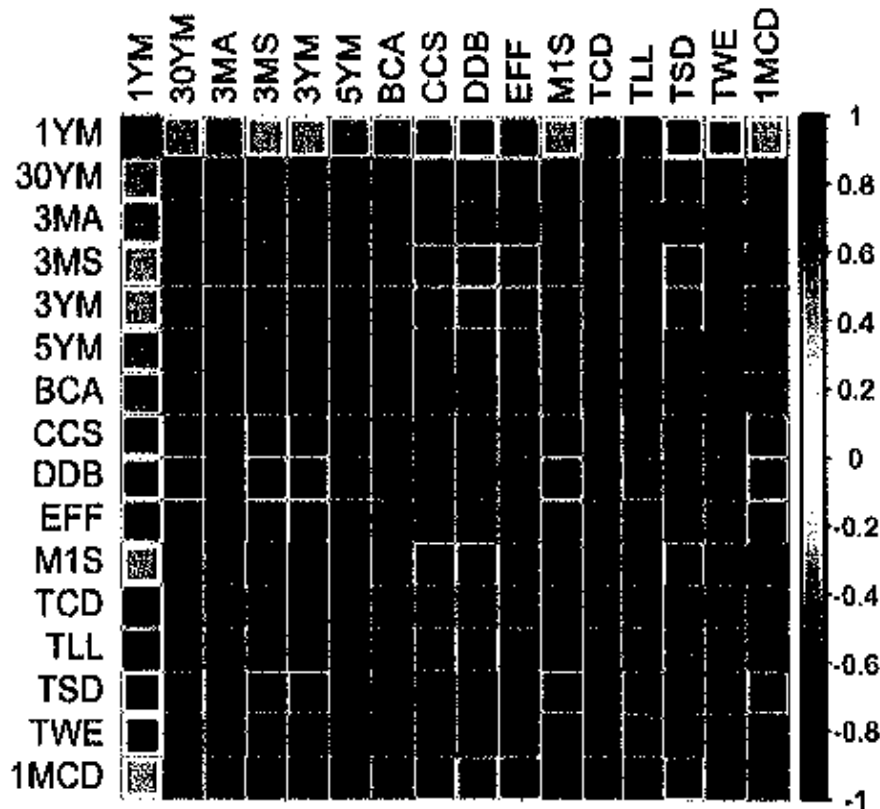


Figure 1: Correlation plot

At the first sight, we notice that our target variables are either positive or negative highly correlated with other explanatory variables except with "1YM" which is nearly half positive correlated. We note that most of the covariables have strong impact on our target variables. However, we have to take into account on correlation of two explanatory variables too. The reason is we want all of our explanatory variables to be independent from each other and if two explanatory variables are highly correlated to each other, they might cause the problem of multicollinearity in our regression model. In another word, if we make a small change in one variable, another one will change too. For example, we see "5YM" and "BCA" are strongly correlated with value 1.0 while "1YM" are weakly correlated to the others.

Federal Reserve Economic Data

1 Présentation de la valeur *Target* et de l'objectif

Le **Certificate of Deposit**, que l'on abrégiera dans la suite en **CD**, est un produit financier particulièrement utilisé aux États-Unis. Il s'agit d'un compte de dépôt comportant deux particularités le différenciant des autres : une période de maturité fixe ainsi qu'un taux d'intérêt fixe. En effet, en échange de taux avantageux et assurés, les clients s'engagent à laisser leur argent dans la banque jusqu'à la fin de la période de maturité, sous peine de pénalités.

Dans notre cas, nous nous intéresserons aux **CD** possédant une période de maturité de 1 mois. Ce type de **CD** est généralement ouvert pour tester la fiabilité de l'organisme financier le proposant, ou pour pouvoir jouir d'une période de maturité minimale lorsque les taux varient énormément selon les mois.

Comme nous pouvons le voir sur la *Figure 1*, les taux de ces **1-Month-CD** peuvent varier drastiquement d'une semaine à l'autre. Il est donc important de pouvoir prédire les futurs taux afin de pouvoir placer son argent dans le timing idéal voir d'augmenter la durée de la période de maturité si une chute de ces derniers est à venir. Pour faire cela, nous nous baserons sur différentes variables traduisant la santé de l'économie américaine.



FIGURE 1 Evolution du taux d'intérêt hebdomadaire des CD de 1 mois entre 1980 et 2000

2 Présentation des variables et de leur lien avec la *Target*

Pour notre étude, nous disposons de différentes variables, établies de façon hebdomadaire entre le 1 Avril 1980 et le 2 Avril 2000 (soit 1049 observations des 16 variables) par la **Federal Reserve Economic Data**.

Dans un premier temps, nous avons accès aux taux de l'échéance constante sur 1 an, 3 ans et 5 ans (respectivement renommés en *TCMR_1*, *TCMR_3* et *TCMR_5*), qui traduisent la valeur fictive du Trésor américain après le laps de temps donné ainsi qu'au taux conventionnel des crédits immobiliers sur 30 ans (*CMR_30*). Nous prenons également en compte le taux appliqué au gouvernement américain pour ses emprunts à court terme sur les marchés primaires et secondaires (respectivement *TBRRA_3* et *TBRSM_3*) ainsi que le taux effectif pratiqué pour les échanges entre établissements financiers (*EFFR*).

Toutes ces valeurs, lorsqu'elles sont élevées, traduisent d'une bonne forme de l'économie américaine, et l'on remarque effectivement dans la *Figure 2* qu'elles sont corrélées positivement avec notre variable cible.

De plus, nous analysons également les valeurs en milliards de dollars des crédits réalisés par les banques américaines (*TTLACB*), de leurs emprunts (*BCACB*), du stock d'argent actuellement disponible (*MMS*) et des demandes de dépôts effectuées (*DDCB*). Parmi ces dépôts, on distingue les dépôts d'épargne (*TSD*) et ceux ouverts par les institutions financières (*TCD*). Avec le pourcentage de liquidité du Trésor américain (*CCMS*), ces variables sont inversement corrélées avec notre variable cible. En effet, contrairement aux variables précédentes, elles traduisent d'une mauvaise forme de l'économie américaine.

Enfin, nous nous baserons également sur un indice traduisant de l'évolution de la valeur du dollar américain à l'international (*TWELMC*), cet indice n'étant a priori pas corrélé avec notre cible.

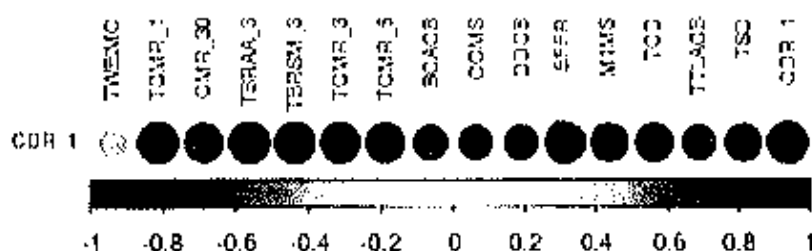


FIGURE 2 Corrélations entre la valeur target et les covariables

PS : Par rapport au sujet proposé, nous avons remarqué une erreur dans la liste des variables données. En effet, les variables 1 à 15 ont toutes été décalées d'un rang (la 1 correspond à la 2, ..., la 15 à la 1).

MRR Project : Statistical Analysis and Description

General nature of the problem:

Financial forecasting has been a challenging area of research since 1980. Modeling financial time series, as the credit of deposit rate, using macroeconomics features is one of the major subjects studied over the last decade explored by several researchers as Anaraki in 2007 in his study of the exchange rate behavior change, as well as Hauner, Lee and Takizawa, 2011 in their study of forecasting financial indexes.

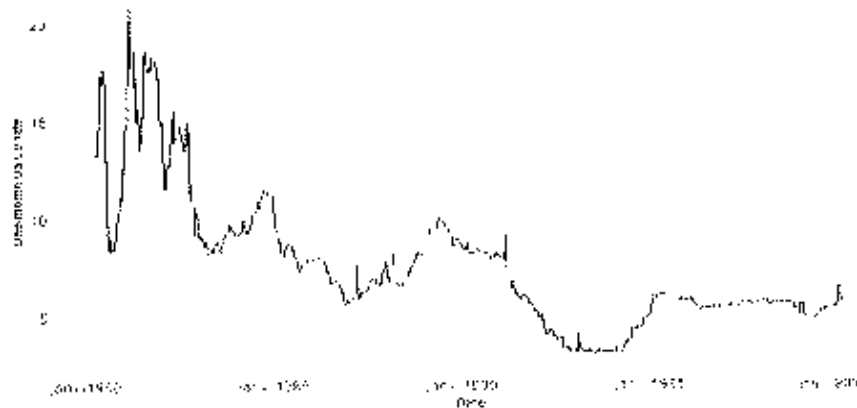
We have been asked to study the Federal Reserve Economic Data in the US. In this data-set our target variable is the One-month certificate of deposit rate (OneMonth CD rate). A certificate of deposit is a time deposit, a financial product commonly sold in the United States by banks. It is characterized by a fixed specific term, in our case it's a one-month term and a fixed interest rate.

Our aim throughout this project, is to predict the one-month CD rate given macroeconomic and financial features of the US market. We will tackle our problem with Penalized regression model (parametric models); Lasso and Ridge; to perform the CD rate forecasting based on economic explanatory variables, as well as K-Nearest-Neighbors: a non-parametric model.

Description of the target variable:

In order to have a proper knowledge of our target variable and its evolution, we plotted in the Graph1 the One-month rate between the dates of 1980/4/1 and 2000/4/2 as it is a continuous variable over time. We can clearly see in this plot that the One-Month rate has a decreasing trend. We notice several peaks on 1981/07/07, as well as 1981/03/15, 1981/08/08, /1982/05/14, the 1984/11/27, also on 1989/06/06 and finally on 1991/03/09. The curve of the CD rate shows that there is no seasonality (no monthly/yearly repetitive pattern). We also thought best to include in the data frame1 the target variable's statistics so as to have a better understanding of its values and characteristics.

Graph1: One-month US CD rate evolution from 1/4/1980 until 2/4/2000



Data frame1 : One-month CD rate statistics

One-month CD rate statistics	
Min.	0.03
1st Qu.	0.57
Median	0.61
Mean	0.52
3rd Qu.	0.82
Max.	20.75

Description of the explanatory variables:

To have a better understanding of the explanatory variables, you will find in the data frame2 below, the names and characteristics of every explanatory variable, its minimum and maximum values as well as its mean value, median value, first and third quartile. We notice here that all the explanatory variables are continuous.

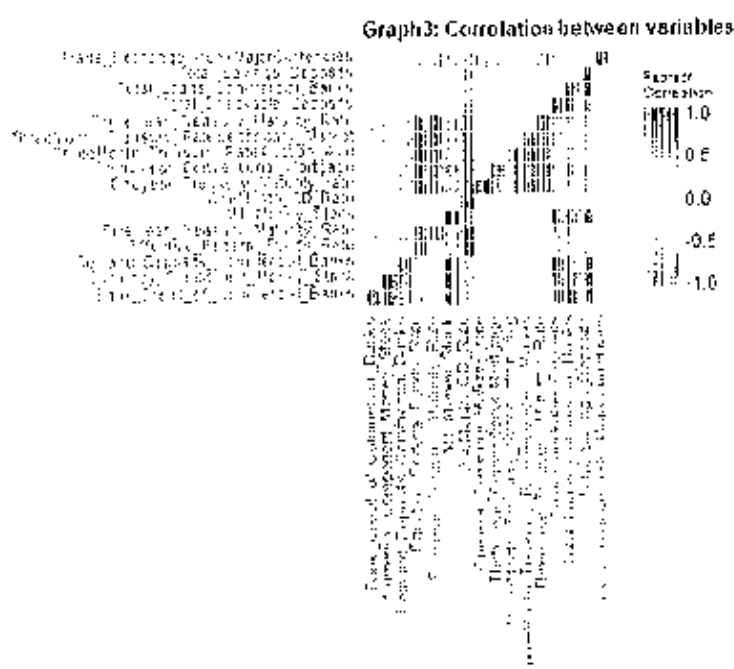
	Trade, Treasury, Policy, Fed, Rate, Vol	Deposit, Treasury, Policy, Fed, Rate, Vol	Trade, Treasury, Policy, Fed, Rate, Vol	Trade, Treasury, Policy, Fed, Rate, Vol	Trade, Treasury, Policy, Fed, Rate, Vol	Trade, Treasury, Policy, Fed, Rate, Vol
1. One-Month	Min: 0.03	Min: 0.03	Min: 0.03	Min: 0.03	Min: 0.03	Min: 0.03
2. One-Month	1st Qu: 0.57	1st Qu: 0.57	1st Qu: 0.57	1st Qu: 0.57	1st Qu: 0.57	1st Qu: 0.57
3. One-Month	Median: 0.61	Median: 0.61	Median: 0.61	Median: 0.61	Median: 0.61	Median: 0.61
4. One-Month	Mean: 0.52	Mean: 0.52	Mean: 0.52	Mean: 0.52	Mean: 0.52	Mean: 0.52
5. One-Month	3rd Qu: 0.82	3rd Qu: 0.82	3rd Qu: 0.82	3rd Qu: 0.82	3rd Qu: 0.82	3rd Qu: 0.82
6. One-Month	Max: 20.75	Max: 20.75	Max: 20.75	Max: 20.75	Max: 20.75	Max: 20.75

Bank Credit of Commercial Banks	Currency Component Money Stock	Demand Deposits Commercial Banks	Effective Federal Funds Rate	One Month CD Rate	Total Checkable Deposits	Total Loans Commercial Banks	Total Savings Deposits
Min: 0.00	Min: 1.00E	Min: 0.00E	Min: 2.000	Min: 0.000	Min: 0.00E	Min: 0.00E	Min: 0.00E
1st Qu: 0.00E	1st Qu: 0.00E	1st Qu: 0.00E	1st Qu: 0.000	1st Qu: 0.000	1st Qu: 0.00E	1st Qu: 0.00E	1st Qu: 0.00E
Median: 0.00E	Median: 0.00E	Median: 0.00E	Median: 0.000	Median: 0.000	Median: 0.00E	Median: 0.00E	Median: 0.00E
Mean: 0.00	Mean: 0.00E	Mean: 0.00E	Mean: 0.000	Mean: 0.000	Mean: 0.00E	Mean: 0.00E	Mean: 0.00E
3rd Qu: 0.00E	3rd Qu: 0.00E	3rd Qu: 0.00E	3rd Qu: 0.000	3rd Qu: 0.000	3rd Qu: 0.00E	3rd Qu: 0.00E	3rd Qu: 0.00E
Max: 0.00	Max: 0.00E	Max: 0.00E	Max: 0.000	Max: 0.000	Max: 0.00E	Max: 0.00E	Max: 0.00E

Data frame2 : Explanatory variables' statistics

The link between the explanatory variables and the target:

To analyze the link between the target variable and the explanatory variables, we plotted the correlation matrix between all variables in the data-set in the Graph3. Pearson correlation seem to be the most relevant correlation coefficient for continuous variables; thus, our matrix correlation plots a 2x2 Pearson correlation. The green color represents negative correlation; the blue color represents positive correlation.



According to the figure, some financial features such as the treasury maturity rate (one, five and three year) seem to be highly and positively correlated to the One Month CD rate. Besides, other economic features like the conventional mortgage and the three month treasury rate seem to be highly and positively correlated with the CD rate. However, the total savings deposits, the total loans to commercial banks and the total checkable deposits seem to be negatively correlated with the CD rate. This might be explained by the fact that the more classical saving deposits customers will have, the lower the credit of deposit rate will be. Besides, the currency component money stock as well as demand deposits at commercial banks and the bank credit of commercial banks are negatively correlated with our target variable. This negative correlation also holds true for the money stock in relation to the One Month CD rate.

Eya Ben Khelifa
César Michelet
Binomial 23

11 novembre 2018

Introduction:

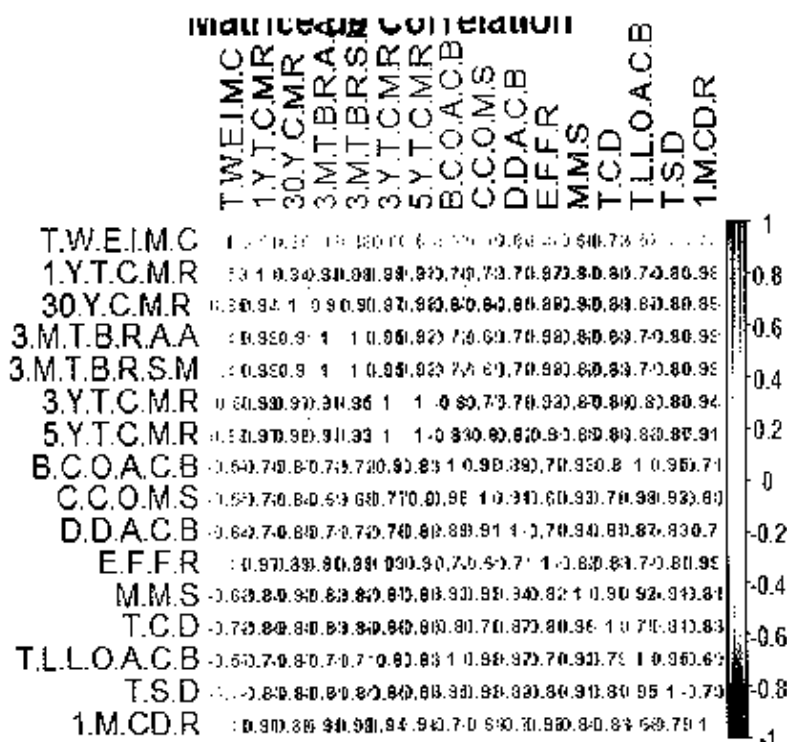
Le monde de la prédiction financière est depuis de nombreuses années un défi majeur dans la data-science. En effet prédire l'évolution de certains produits financiers ou les futures chiffres d'une banque ou d'un groupe permettrait, au-delà de s'enrichir pour certains, de prévoir une quelconque crise et ainsi atteindre une certaine stabilité économique au sein d'une entreprise ou bien même d'un pays.

Ainsi notre projet consiste donc à évaluer et prédire une variable à travers un jeu de données incluant cette variable et d'autres variables influant l'évolution de la variable à prédire.

Data:

On dispose de données hebdomadaires concernant plusieurs variables entre le 01/04/1980 et 02/04/2000. Le but est de prédire la variable "taux de CD d'un mois" qui est la 16^{ème} variable.

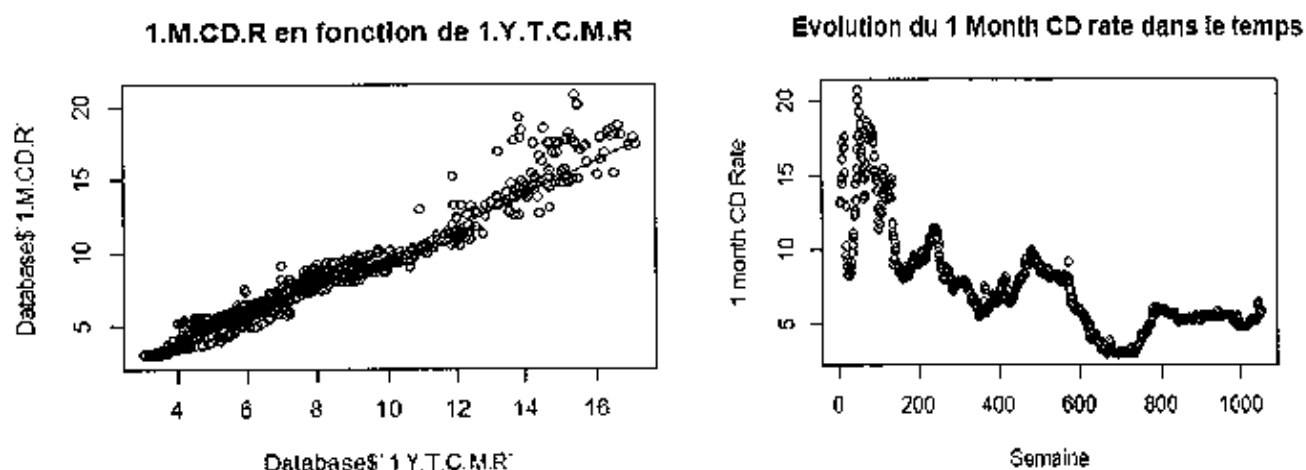
Matrice de corrélation:



Après avoir réalisé cette matrice de corrélation, on constate qu'il y a des variables fortement corrélées avec la variable à prédire (la 1 Month CD Rate) dont la 2ème, 3ème, 4ème, 5ème, 6ème, 11ème, 12ème et 13ème variables. Ces variables corrélées à la variable à prédire et sont fortement corrélées entre elles même puisque la 2ème variable (1-year Treasury Constant Maturity rate) est fortement corrélée aux autres, donc on va tracer le graphe liant notre variable à prédire et la 2ème variable et qui va ressembler aussi au graphe entre la variable à prédire et les autres variables corrélées à elle.

Il y a des variables dont le coefficient de corrélation est égal à 1 donc à priori ces variables comportent les mêmes informations, c'est le cas pour la variable "3-Month Treasury Bill Rate-Auction Average" et la variable "3-Month Treasury Bill Rate-Secondary Market", la variable "3-Year Treasury Constant Maturity Rate" et la variable "5-Year Treasury Constant Maturity Rate", la variable "Bank Credit of All Commercial Banks" et la variable "Total Loans and Leases of All Commercial Banks", ce qui est totalement logique puisque ces variables sont liées du point de vue financier.

Relation entre la variable à prédire et le 1-Year Treasury Constant Maturity Rate et évolution de la variable à prédire dans le temps :



Le premier graphe montre l'évolution de la variable à prédire (1 month CD rate) en fonction de la 1.Y.T.C.M.R.

Le 2ème montre l'évolution du 1.M.CD.R en fonction de l'indice de la semaine d'étude et on remarque que notre variable est instable, ses valeurs changent beaucoup, ce qui donne un sens à notre étude prédictive.

Projet de MRR - Détails du sujet

« Federal Reserve Economic Data »

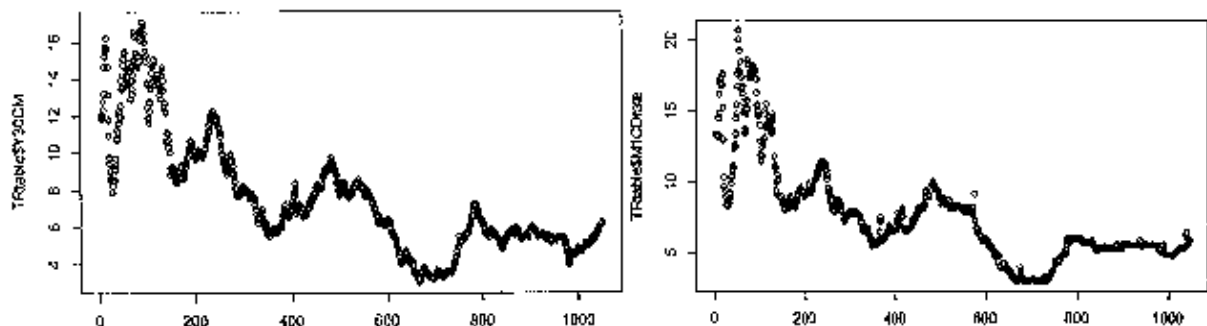
Le dataset provient de la banque de réserve fédérale de St-Louis, aux États-Unis. Elle recense 20 années de statistiques économiques américaines, à partir du 01/04/1980 jusqu'au 02/04/2000.

L'objectif est ici de prédire le « 1-Month CD rate », c'est-à-dire le taux d'intérêt des certificats de dépôt pour une durée d'un mois. Un certificat de dépôt est une somme déposée à la banque sur laquelle on perçoit des intérêts au retrait.

15 autres variables sont disponibles pour éventuellement aider à prédire la tendance du taux d'intérêt d'un mois après Avril 2000 :

- 1-year Treasury Constant Maturity rate
- 30-Year Conventional Mortgage Rate
- 3-Month Treasury Bill Rate-Auction Average
- 3-Month Treasury Bill Rate-Secondary Market
- 3-Year Treasury Constant Maturity Rate
- 5-Year Treasury Constant Maturity Rate
- Bank Credit of All Commercial Banks
- Currency Component of Money Stock
- Demand Deposits at Commercial Banks
- Effective Federal Funds Rate
- M1 Money Stock
- Total Checkable Deposits
- Total Loans and Leases of All Commercial Banks
- Total Savings Deposits
- Trade Weighted Exchange Index Major Currencies

Bien que la plupart des variables soient des taux d'intérêt/prêt, on peut observer qu'ils partagent des tendances similaires à l'échelle des 20 années d'étude, comme on peut le voir ci-dessous, indiquant un potentiel lien entre notre variable cible et les covariables.



A gauche : la courbe du taux d'intérêt immobilier sur 30 ans ; à droite l'évolution du taux d'intérêt des certificats de dépôt sur 1 mois

