Yield curve inversion: An early warning sign of financial crises

Léo Colin, Alexis Christien, Saikou Sounounou Bah, Louis Lebreton

Directed by Philippe De Peretti

Master 1 Econometrics-Statistics 2023-2024

Abstract

In 1995, Arturo Estrella and Frederic S. Mishkin highlighted a link between the inversion of the yield curve and a future economic recession within 6 to 18 months. Their work was subsequently replicated with very convincing results. This thesis aims to replicate the studies of Mishkin and Estrella and to validate their study across multiple countries over the period from 1986 to 2024 using OECD data.

Contents

1 Introduction

2 Context

To better understand our study and the one conducted by Estrella and Mishkin, it is important to provide some clarification regarding the concepts studied.

Mishkin and Estrella's study focuses on U.S. Treasury bonds. These are government-backed debt securities issued by the US Treasury. But the study can be carried out on other countries that have their own bonds. Investors who buy these securities receive interest rates in return, depending on their maturity. These are attractive investments, as they are recognized as low-risk, with high liquidity and flexible time horizons. An important link is the inverse relationship between security prices and yields. Therfore, the higher the yield, the lower the price.

Yield curves measures the interest rates depending on maturity. Here's a simplified representation of a normal yield curve :

Figure 1: Normal Yield Curve

With :

- X-axis: The maturity of the bonds
- Y-axis : yield received annually

Generally speaking, the curve slopes upwards because when you lend money to the government, you are taking a greater risk by leaving it at the government's disposal for a longer period, and you therefore want a higher interest rate. This means that a 20-year bond generally yields more than a 3-year bond.
The authors identify this inversion as a signal of a future recession. But why can this curve invert? This is due to two major reasons, each affecting one side of the curve. Before explaining them, it is essential to understand that the inversion of the curve is part of economic cycles. Economies are composed of cycles, periods of economic growth followed by periods of recession and so on. To understand the reasons for the inversion, let's suppose a context where an economy is not yet in recession, that is, in a period of growth.

- The first reason is **central banks**. They aim to avoid extreme economic phenomena by playing with the instruments at their disposal, including short-term interest rates on bonds. In a period of economic growth, the central bank wants to avoid an overheating of the economy that could quickly turn into uncontrollable inflation. To avoid this, it increases these short-term interest rates. This reduces investments and loans and slows down the economy. In this case, we understand that the left side of the curve will increase.
- The second reason, which explains the right-hand side of the curve, comes from **investors**. When they foresee a downturn in the economy, they turn to safe, long-term assets such as bonds. This increase in long-term bonds demand will lead to higher prices and lower yields. The right-hand side of the curve then declines.

If we combine these two effects, we arrive at an inverted yield curve :

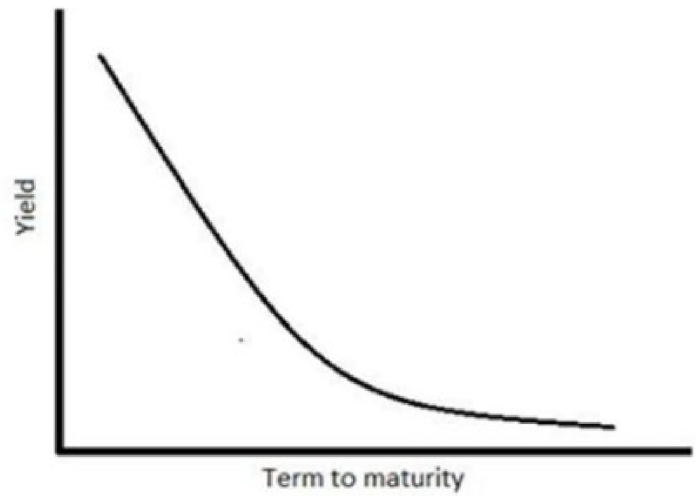

Figure 2: Inverted Yield Curve

3 The paper of Estrella and Mishkin

3.1 Abstract

The article "Predicting US Crises" by Estrella and Mishkin (1995) examines the use of the yield curve as an indicator of economic recessions. The authors demonstrate that the difference between short-term and long-term interest rates (often referred to as the slope of the yield curve) is a reliable predictor of recessions in the United States. In fact, it has been observed that the inversion of the yield curve has preceded every recession in the United States since the 1960s.

3.2 Their research

By analyzing historical data, they show that when an inversion of the yield curve occurs, there is a high probability of a recession following within 12 to 18 months. To demonstrate this, they focus on predicting a recession using the spread in percentage points between rates of different maturities. The advantage of using a single financial indicator is to avoid overfitting, thereby simplifying the model and making it more robust.

The model they propose is as follows:

$$P(R_{t+k} = 1) = F(a_1 + a_2 X_{1t} + a_3 X_{2t} + \ldots)$$

where

$$P(R_{t+k} = 1) = \begin{cases} 1 & \text{if quarter } t+k \text{ is a recession} \\ 0 & \text{otherwise} \end{cases}$$

They use a probit model to predict the probability of a recession, where F is the normal cumulative distribution function. This model provides a probability close to 1 for a strong prediction of a recession, and close to 0 for the opposite.

For example, one of the most successful models in the paper is the probit equation for predicting a recession four quarters ahead, using only the spread between the 10-year Treasury note and the 3-month Treasury bill (SPREAD) as the explanatory variable. The estimates, using data from the first quarter of 1960 to the first quarter of 1995, show the following results:

$$P(R_{t+4} = 1) = F(-0.66 - 0.81\mathrm{SP}_t)$$

- When the spread averages 0.76 percentage points over the quarter, the estimated probability of a recession four quarters ahead is 10%.
- When the spread averages -0.82 percentage points, the probability is 50%.
- When the spread averages -2.40 percentage points, the probability is 90%.

These findings illustrate the strong predictive power of the yield curve slope in forecasting economic downturns, highlighting its importance as a financial indicator.

For the Eurozone, a study conducted by Sabes and Sahuc explores how yield curve inversions can predict recessions. They base their analysis on a period spanning from 1970 to 2022 and use the AUROC (Area Under the Receiver Operating Characteristic) metric to evaluate the predictive effectiveness of the yield curves. The results show that the yield curve is a good predictor of recession at the aggregated Eurozone level, with an AUROC of 0.92 for the period 1970-2008. However, this performance decreased after the global financial crisis, with the AUROC falling to 0.75 for the period 1970-2022. The unconventional monetary policies implemented by the ECB and the unique characteristics of recent recessions, such as the one caused by the COVID-19 pandemic, are factors that have blurred this signal.

The study also analyzes individual countries. The yield curves of France and Germany show good predictive performance, with AUROCs exceeding the threshold of 0.7. In Germany, the data indicates that the yield curve has reliably predicted recessions, with an AUROC of 0.84 for the period 1970-2008. This is partly explained by Germany's relative economic stability and the perception of its sovereign credit as very safe. In contrast, for peripheral countries (Italy, Spain), the relationship is less clear, with AUROCs not exceeding 0.7.

According to the Bulletin of the Banque de France (January-February 2024), the yield curve inversion observed in June 2023 in the Eurozone has raised fears of a recession. Despite the negative slope of the yield curve, the ECB's December 2023 projections estimate GDP growth of 0.8 percent for the Eurozone in 2024. The study also highlights that while the yield curve inversion has been a reliable indicator for predicting past recessions, its effectiveness has recently diminished due to unconventional monetary policies and specific economic circumstances.

4 Simple logit model

4.1 Creation of the data frame

To create our dataframe to work on, we used OECD data, which provide short-term and long-term rates as well as quarterly growth rates. We created a new variable, the spread, calculated as the difference between the long-term rate and the short-term rate (i long term - i short term).A quarter is considered to be in recession when the quarterly growth rate is negative and the previous quarter's growth rate is also negative. This allows us to define a binary variable "recession," which takes the value 1 in case of a recession and 0 otherwise. Next, we created a "future recession" variable that is equal to 1 if there is at least one quarterly recession between t+2 and t+6 (corresponding to 6 to 18 months, in accordance with the study by

Mishkin and Estrella), and 0 otherwise.

Obs	IRLT	GDP_Growth_Rate	IRST	spread	Recession	Recession_future
1	9.8040	-0.0507	10.9700	-1.16600	1	1
2	8.9668	0.5748	8.9417	0.02517	0	1
3	8.8998	0.1257	8.4442	0.45567	0	0
4	8.9115	-0.7239	8.3958	0.51567	1	0
5	8.3567	2.3114	7.5125	0.84417	0	0
6	9.2510	1.2810	8.1108	1.14017	0	0
7	10.1303	1.5279	9.0458	1.08450	0	0
8	10.1448	1.2919	8.9692	1.17567	0	0
9	9.4457	1.4570	8.6175	0.82817	0	0

Figure 3: Built data frame

4.2 Logit model results ”In sample”

Let’s start by estimating the logistic model in-sample, that is, with our entire database. With a simple logistic model, we obtain the following results:

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	31.9293	1	<.0001
Score	30.9716	1	<.0001
Wald	29.3776	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.1641	0.0836	3.8499	0.0497
spread	1	-0.2559	0.0472	29.3776	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
spread	0.774	0.706	0.849

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	61.9	Somers' D	0.238
Percent Discordant	38.1	Gamma	0.238
Percent Tied	0.0	Tau-a	0.119
Pairs	193997	c	0.619

Figure 4: Odds ratio and parameters obtained (In sample)

Thanks to these results, we can say that each additional percentage point of spread between the long term rate and the short term rate significantly reduces (at the 1% threshold) the expectation of the probability that at least one quarterly recession will appear over the next 6 to 18 months by 22.6%. This seems to confirm the theory of Mishkin and Estrella.

By choosing a classification threshold of 0.5, we obtain the following confusion matrix:

Table of Recession_future by predicted_class			
	predicted_class		
Recession_future	0	1	Total
0	334	129	463
1	215	204	419
Total	549	333	882

Figure 5: Confusion matrix (In sample)

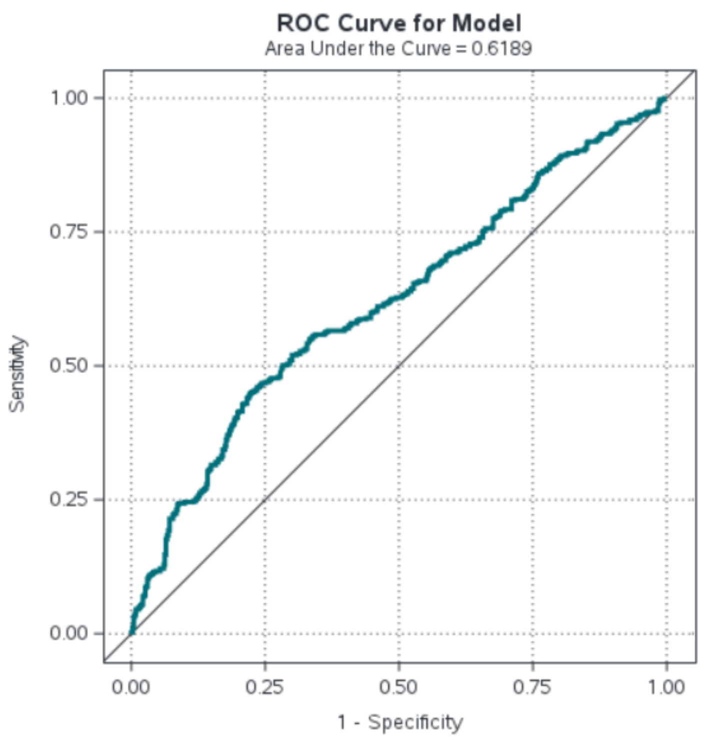

Figure 6: ROC curve and AUC (In sample)

The AUC is 0.61, so our model predicts only slightly better than chance.

4.3 Logit model results "Out of sample"

Now we implement our model out-of-sample with a training dataset that represents 80% of the total database. With a simple logistic model, we obtain the following results:

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	24.4854	1	<.0001
Score	23.7981	1	<.0001
Wald	22.5758	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.2171	0.0927	5.4849	0.0192
spread	1	-0.2462	0.0518	22.5758	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
spread	0.782	0.706	0.865

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	61.5	Somers' D	0.229
Percent Discordant	38.5	Gamma	0.229
Percent Tied	0.0	Tau-a	0.115
Pairs	124573	c	0.615

Figure 7: Odds ratio and parameters obtained (Out of sample)

Here, we can say that each additional percentage point of spread between the long term rate and the short term rate significantly reduces (at the 1% threshold) the expectation of the probability that at least one quarterly recession will appear over the next 6 to 18 months by 21.8%. Once again the theory of Mishkin and Estrella is confirmed.

By choosing a classification threshold of 0.5, we obtain the following confusion matrix:

Table of Recession_future by predicted_class			
	predicted_class		
Recession_future	0	1	Total
0	51	2	53
1	23	0	23
Total	74	2	76

Figure 8: Confusion matrix (Out of sample)

As the recession modality is poorly represented, it is mainly false negatives that we obtain. Our model never succeeded in predicting a crisis when there was one. Perhaps we should add more variables to make our model more discriminating.

5 Logit model with more explanatory variables

5.1 New data frame and explanatory variables added

We have added the following variables: Inflation (percentage per quarter), the increase in the M3 monetary aggregate (percentage per quarter), and the increase in unemployment (percentage per quarter). We kept the same countries but had to start from 1991 instead of 1986 to have complete data, as we did not have unemployment data for Germany prior to the reunification of the country.

It is plausible to believe that these three variables could influence the probability of a recession within a 6 to 18-month horizon. Rising inflation can erode consumer purchasing power and increase costs for businesses, potentially slowing economic growth. An increase in the M3 monetary aggregate may signal excessive money supply, which can lead to inflationary pressures and economic instability. It can also be attributed to expansionary monetary policy, which aims to stimulate the economy during a slowdown. Lastly, rising unemployment can reduce consumer spending and confidence, which are critical drivers of economic activity. Therefore, monitoring these variables can provide valuable insights into the likelihood of an impending recession.

5.2 Logit model results with more explanatory variables "In sample"

Let's start by estimating the logistic model in-sample, that is, with our entire database. We obtain the following results:

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	478.277	409.439
SC	482.220	429.153
-2 Log L	476.277	399.439

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	76.8382	4	<.0001
Score	67.2333	4	<.0001
Wald	55.9309	4	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.8229	0.3590	5.2542	0.0219
spread	1	-0.9441	0.1441	42.9223	<.0001
Inflation	1	0.00718	0.0923	0.0060	0.9380
M3_Growth_Rate	1	-0.3712	0.0615	36.3958	<.0001
Unemployment_Growth_	1	0.5958	0.1934	9.4877	0.0021

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
spread	0.389	0.293	0.516
Inflation	1.007	0.840	1.207
M3_Growth_Rate	0.690	0.612	0.778
Unemployment_Growth_	1.814	1.242	2.651

Figure 9: Odds ratio and parameters obtained (In sample)

Thanks to these results, we can say that each additional percentage point of spread between the long term rate and the short term rate significantly reduces (at the 1% threshold) the expectation of the probability that at least one quarterly recession will appear over the next 6 to 18 months by 71.1% all our variables being equal. This confirms the theory of Mishkin and Estrella.

By choosing a classification threshold of 0.5, we obtain the following confusion matrix:

Table of Recession_future by predicted_class			
	predicted_class		
Recession_future	0	1	Total
0	229	31	260
1	66	55	121
Total	295	86	381

Figure 10: Confusion matrix (In sample)

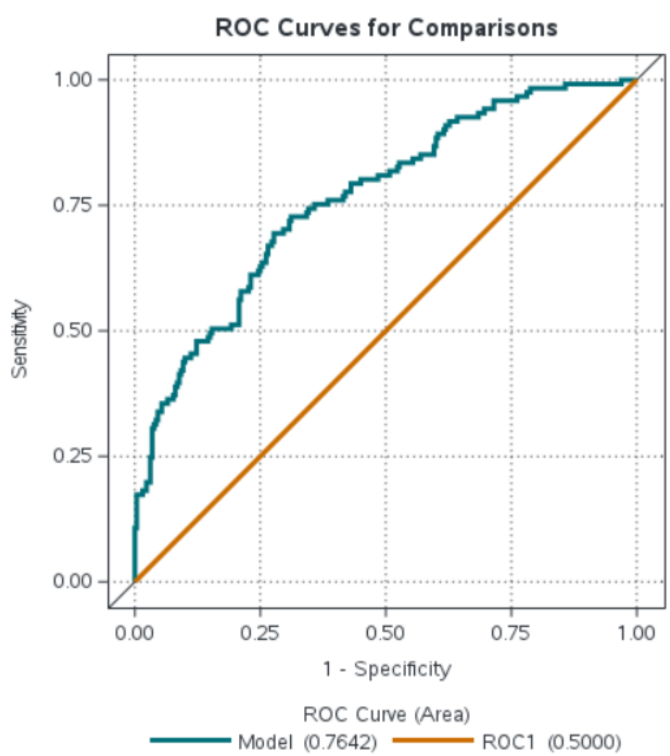

Figure 11: ROC curve and AUC (In sample)

The AUC is 0.76, so our model predicts quite correctly.

5.3 Logit model results with more explanatory variables results "Out of sample"

Now we implement our model out-of-sample with a training dataset that represents 80% of the total database. With our logistic model, we obtain the following results:

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	384.991	328.504
SC	388.711	347.106
-2 Log L	382.991	318.504

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	64.4867	4	<.0001
Score	56.7911	4	<.0001
Wald	46.8967	4	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.8044	0.3871	4.3185	0.0377
spread	1	-0.9708	0.1624	35.7320	<.0001
Inflation	1	-0.0473	0.0993	0.2266	0.6340
M3_Growth_Rate	1	-0.3571	0.0664	28.9452	<.0001
Unemployment_Growth_	1	0.7384	0.2119	12.1438	0.0005

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
spread	0.379	0.276	0.521
Inflation	0.954	0.785	1.159
M3_Growth_Rate	0.700	0.614	0.797
Unemployment_Growth_	2.093	1.381	3.170

Figure 12: Odds ratio and parameters obtained (Out of sample)

By choosing a classification threshold of 0.5, we obtain the following confusion matrix:

Table of Recession_future by predicted_class			
	predicted_class		
Recession_future	0	1	Total
0	54	1	55
1	15	6	21
Total	69	7	76

Figure 13: Confusion matrix (Out of sample)

Our model now predicts better than before but it still has difficulty predicting crises when there is one.

5.4 LASSO logistic regression

5.4.1 Principles

Here, we have several explanatory variables so we can integrate a LASSO penalty into the cost function of our logistic model. LASSO is used in logistic regression for automatic variable selection and regularization, which simplifies the model by setting some coefficients to zero and reducing multicollinearity. This improves prediction accuracy and prevents overfitting, especially with high-dimensional data. As a result, the model becomes more interpretable and computationally efficient.

The cost function for logistic regression with LASSO (Least Absolute Shrinkage and Selection Operator) regularization combines the negative log-likelihood of logistic regression with an L1 penalty. Here are the components of this cost function:

1. **Negative log-likelihood**: This part measures the prediction error of the logistic regression. 2. **L1 Penalty**: This part penalizes the sum of the absolute values of the regression coefficients to encourage sparsity, meaning that some coefficients will be exactly zero.

The cost function for logistic regression with LASSO can be formulated as follows:

$$J(\beta) = -\left[\frac{1}{N}\sum_{i=1}^{N}(y_i \log(h_\beta(x_i)) + (1-y_i)\log(1-h_\beta(x_i)))\right] + \lambda\sum_{j=1}^{p}|\beta_j|$$

where:

- N is the number of observations.
- y_i is the observed value of the dependent variable for observation i.
- x_i is the vector of explanatory variables for observation i.
- $h_\beta(x_i)$ is the predicted probability that $y_i = 1$ given by the logistic model:
$$h_\beta(x_i) = \frac{1}{1+\exp(-\beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})}$$
- β is the vector of regression coefficients (including the intercept β_0).
- λ is the regularization parameter that controls the importance of the L1 penalty.

5.4.2 Results

Fit Statistics	
-2 Log Likelihood	399.49
AIC (smaller is better)	409.49
AICC (smaller is better)	409.65
BIC (smaller is better)	429.21
Pearson Chi-Square	380.72
Pearson Chi-Square/DF	1.0126

Parameter Estimates		
Parameter	DF	Estimate
Intercept	1	0.808297
spread	1	-0.922452
Inflation	1	0.002862
M3_Growth_Rate	1	-0.362249
Unemployment_Growth_	1	0.562307

Figure 14: Lasso logistic regression results

Table of Recession_future by predicted_class_lasso

	predicted_class_lasso		
Recession_future	**0**	**1**	**Total**
0	231	29	260
1	66	55	121
Total	297	84	381

Figure 15: Confusion matrix (Lasso - In sample)

We obtain almost similar prediction performances with and without the LASSO penalization.

Conclusion

In conclusion, our research and analyses confirm the work of Mishkin and Estrella. The inversion of the yield curve proves to be a reliable indicator of economic recessions, typically occurring within a horizon of 6 to 18 months after the inversion. The historical data and economic models we have examined support the idea that this leading indicator deserves close attention from economists, policymakers, and investors. Therefore, monitoring the inversion of the yield curve can provide valuable warnings and enable proactive measures to mitigate the potential impacts of an impending recession.

References

Estrella, Arturo and Frederic S. Mishkin. 1995. "Predicting US recessions: financial variables as leading indicators." *NBER Working Papers 5379* .

6 Appendix - Our main SAS code

```
/*  data import */
proc import datafile="/shared/home/Louis.Lebreton@etu.
   univ-paris1.fr/casuser/
   data_interest_rates_more_variables.csv"
            out=interest_rates
            dbms=csv
            replace;
    getnames=yes;
run;

/* Re-organizing the dataframe */
data interest_rates2;
        set interest_rates(keep= MEASURE OBS_VALUE
           Reference_area time_period);
run;

proc sort data=interest_rates2;
        by reference_area time_period;
run;

proc transpose data=interest_rates2 out=interest_rates3;
        by reference_area time_period;
        id measure;
        var obs_value;
run;


data interest_rates3(drop=_name_);
        set interest_rates3(rename=(B1GQ=GDP_Growth_Rate
            UNEM=Unemployment_Growth_Rate IR3T=IRST
           reference_area=Country CP=Inflation MABM=
           M3_Growth_Rate));
run;


/* new variable spread : i long term - i short term */
data interest_rates3;
```

```
    set interest_rates3;
    spread = IRLT - IRST;
run;

data interest_rates3;
    set interest_rates3;
    if GDP_Growth_Rate < 0 then Recession = 1;
    else Recession = 0;
run;

/* new variable Recession_future :  Recession between t
   +2 to t+6 */
proc iml;
    /* convert to matrix */
    use interest_rates3;
    read all var {Recession} into Recession_vector;
    close interest_rates3;

    /* new columns */
    n = nrow(Recession_vector);
    Recession_sum = j(n, 1, .);
    Recession_future = j(n, 1, .);

    /* filling of Recession_sum */
    do i = 1 to (n-6);
        Recession_sum[i] = Recession_vector[i+2] +
           Recession_vector[i+3] + Recession_vector[i+4]
            + Recession_vector[i+5] + Recession_vector[i
           +6];
    end;

    /* filling of Recession_future */
    do i = 1 to n;
        if Recession_sum[i] = 0 then
            Recession_future[i] = 0;
        else
            Recession_future[i] = 1;
    end;

    /* correction of irrelevant values at the end for
```

```
   each country */
/* number of quarters by country */
    n_q= (2024-1990)*4 - 3;
    do i = 1 to n;
    if mod(i, n_q)=0 then do;
                    do j=0 to 5;
            Recession_future[i-j]=.;
                    end;
            end;

end;

/* adding Recession_future to the dataframe */
use interest_rates3;
read all into interest_rates3_data;
close interest_rates3;

interest_rates3_data = interest_rates3_data ||
   Recession_future;

create interest_rates4 from interest_rates3_data[
   colname={"IRLT" "GDP_Growth_Rate" "IRST" "
   Unemployment_Growth_Rate" "Inflation" "
   M3_Growth_Rate" "spread" "Recession" "
   Recession_future"}];
append from interest_rates3_data;
close interest_rates4;
quit;


/* Removal of row with missing (incalculable) values */
data interest_rates4;
    set interest_rates4;
    if cmiss(of _all_) then delete;
run;

/* Final dataframe */
proc print data=interest_rates4;
run;
```

```
/* Stats */
proc contents data=interest_rates4;
run;

proc means data=interest_rates4;
        var IRST IRLT GDP_Growth_Rate SPREAD RECESSION
           RECESSION_Future;
run;
/* we can see that the results are consistent */

proc univariate data=interest_rates4;
        var IRST IRLT GDP_Growth_Rate Inflation
           M3_Growth_Rate Unemployment_Growth_Rate
           SPREAD RECESSION RECESSION_Future;
run;

/* OUT OF SAMPLE */
/* data train test separation */

/*random selection*/
proc surveyselect data=interest_rates4 rate=0.8 outall
   out=selection seed=111;
run;

/*df train and test*/
data df_train df_test;
set selection;
if selected =1 then output df_train;
else output df_test;
drop selected;
run;

/*logistic model*/
proc logistic data=df_train;
    model recession_future(event='1') = spread Inflation
        M3_Growth_Rate Unemployment_Growth_Rate;
    store out=logistic_model;
run;
```

```
/*prediction on df test*/
proc plm restore=logistic_model;
    score data=df_test out=pred_test predicted=prob;
run;


data pred_test;
    set pred_test;
        /*classification threshold here : 0.5 */
    predicted_class = (prob >= 0.5);
run;

/* confusion matrix */
proc freq data=pred_test;
    tables recession_future*predicted_class  / norow
       nocol nopercent;
run;

/*IN SAMPLE*/
/* logit model in sample : data=interest_rates4
or out of sample : data=df_train*/

proc logistic data=interest_rates4;
    model recession_future(event='1') = spread Inflation
        M3_Growth_Rate Unemployment_Growth_Rate ;
    output out=pred p=prob;
run;

data pred;
    set pred;
        /*classification threshold here : 0.5 */
    predicted_class = (prob >= 0.5);
run;

/* confusion matrix */
proc freq data=pred;
    tables recession_future*predicted_class  / norow
       nocol nopercent;
run;
```

```
/* AUC */
proc logistic data=interest_rates4;
    model recession_future(event='1') = spread Inflation
        M3_Growth_Rate Unemployment_Growth_Rate ;
    roc;
run;

/* logit model with LASSO penalization */
proc hpgenselect data=interest_rates4;
    model recession_future(event='1') = spread Inflation
        M3_Growth_Rate Unemployment_Growth_Rate
       Inflation M3_Growth_Rate Unemployment_Growth_Rate
        / dist=binomial link=logit;
    selection method=lasso;
    output out=pred_lasso p=prob_lasso;
run;
data pred_lasso;
    set pred_lasso;
        /*classification threshold here : 0.5 */
    predicted_class_lasso = (prob_lasso >= 0.5);
run;
/* confusion matrix lasso*/
data pred_lasso;
    set pred_lasso;
    ID = _N_;
run;

data pred;
    set pred;
    ID = _N_;
run;
data pred_lasso_nolasso;
   merge pred pred_lasso;
   by ID;
run;

proc freq data=pred_lasso_nolasso;
    tables recession_future*predicted_class_lasso  /
       norow nocol nopercent;
run;
```