ISSN 2063-5346

FINANCIAL CRISIS PREDICTION USING MACHINE LEARNING TOOLS - A CASE STUDY ON UK PENSION CRISIS

Dr. Vanisree Talluri[1], Ms. V. Vidya Lakshmi[2]

Article History: Received: 10.05.2023 Revised: 29.05.2023 Accepted: 09.06.2023

Abstract

Financial crises are a deterrent to economic growth. It is important to understand past patterns and trends to predict their occurrence. The focus of this study is to understand how these patterns affect the prediction of financial crises so that an economy can safeguard itself from its consequences. This study considers four predictors namely, credit growth, public debt, household (HH) final consumption and Gross Domestic Product (GDP). For the sake of convenience credit growth and public debt are taken as a ratio of GDP and then the change in the respective ratios over several quarters has been analysed. Quarterly growth of household final consumption is analysed on Year-on-Year basis. The sample consists of data between the years 2018 and 2022. Once the pattern using actual data is established, governments can apply machine learning tools on their estimates to see how long it may continue and make necessary provisions to safeguard the economy. The study deploys machine learning tools such as logistic regression to predict the probability of the UK Pension Crisis which began in September 2022.

Key words: Financial crises, GDP, Public debt, HH final consumption, UK Pension Crisis, Logistic Regression, Machine Learning Tools.

[1]Asst. Professor, St. Francis College for Women, Hyderabad

[2]M.Com Student, St.Francis College for Women, Hyderabad

DOI:10.48047/ecb/2023.12.9.39

1.1 INTRODUCTION

"As sure as the spring will follow winter, prosperity and economic growth will follow recession" - Robert Foster Bennet

The UK economy has been one of the most watched economies ever since it announced Brexit. Having said that, it has faced several issues which were later followed by the global pandemic. The recent mini budget announced by the former UK Finance Minister not only shook the economy but also changed peoples' perception about government issued instruments. These events took a toll on the UK economy resulting in high inflation rates and slipping into global recession faster than other countries.

Predicting financial crises' occurrence can help economies fight them and make necessary arrangements to allocate resources. A financial crisis may have several causes. In the case of the UK, the cause can be narrowed down to the herd mentality of investors who wanted to get rid of gilts as the yields were rising. This not only triggered the financial crisis, but also forced the Bank of England to make a forced intervention and assay the role of buyer of last resort.

This study developed a logistic regression model and applied it on four variables. This was used on actual data for the years 2018-2022 and it came up with almost 99% probability of occurrence of the UK Pension Crisis. Further, the same model can be applied on estimates of coming years to forecast how long the current crisis will continue and if used consistently in future years, it will help economies predict any impending financial crises.

1.2 REVIEW OF LITERATURE

By using machine learning techniques to analyse macro-financial data for 17 nations from 1870 to 2016, Kristina Bluwstein[1](2021) created early warning models for financial crisis prediction.

The slope of the yield curve on a global and domestic level, as well as credit expansion, are the two most significant determinants throughout. This research attempts to forecast financial crises up to two years in advance using a varied mix of models from the machine learning field. Should models indicate a high likelihood of a crisis, this would give decision-makers time to respond, for example by activating macroprudential regulations like countercyclical capital buffers. A toolkit of adaptable models is provided by machine learning.

They have been demonstrated to be more accurate than benchmark traditional econometric models in many prediction tasks, and they are novel in applications linked to economics, especially in cases where many different factors play a role, and the relationship between these factors is complex.

Using historical data on post-war financial crises worldwide, Robin Greenwood[2](2021) demonstrates that crises are largely predictable. There is a 40% chance of experiencing a financial crisis during the next three years if fast loan expansion and asset price growth have been present over the previous three years in either the nonfinancial company or household sectors. In normal circumstances, when neither credit nor asset price growth has been elevated, there is a probability of about 7%. The evidence refutes the idea that financial crises are unforeseen "bolts from the sky" and supports the hypothesis put out by Kindleberger and Minsky that crises are a by-product of predictable, boom-bust credit cycles.

1.3 STATEMENT OF PROBLEM:

Prediction of a financial crisis has become important to understand the onset of recessions and the effect it has on the economy in the long run. The study takes the data of the UK Pension Crisis and tries to understand what happened to the markets on the day of the crash, and using the logistic regression model, how the government can make future predictions and allocate resources.

Therefore, the statement of problem is:

"Predicting the probability of financial crisis using macroeconomic predictors with the help of machine learning tools."

1.4 OBJECTIVE

- To apply a machine learning tool on macroeconomic factors to predict the probability of occurrence of financial crisis.

1.5 RESEARCH METHODOLOGY

The following methodology has been used for Analysis and Interpretation:

1.5.1 Data Collection:

To fulfil the objectives of the study, only *secondary* data has been used. These are published reports on the official Bank of England and Office for National Statistics, UK websites.

Few other *Secondary data statistics* for the study are collected from published records, websites, and reports available online. The data required for this study is related to Economics with a few extra calculations required to convert monthly data to quarterly data and as a percentage of Gross Domestic Product (GDP), wherever required.

1.5.2 Period of Study

The study is limited to a period of four years, that is 2018 to 2022.

1.5.3 Tools Used

Logistic regression has been used to analyze data using Microsoft Excel. The computerized package used for this purpose is Excel Solver.

1.6 DATA ANALYSIS

The collected and computed data have been analyzed through the machine learning tool of logistic regression and the results are as follows:

Table 1 – The Initial Input Data

QUARTERS	HH FINAL CONUMPTION - QUARTERLY GROWTH RATE OF THE INDEX (YOY)	QUARTERLY DIFFERENCE OF GDP RATIO*100 - CREDIT GROWTH	QUARTERLY DIFFERENCE OF GDP RATIO*100- PUBLIC DEBT	QUARTERLY GDP GROWTH (YOY)	ACTUAL OCCURANCE OF FINANCIAL CRISIS
2018 Q1	1.923071181	-56.78579378	-1.3221154	3.4	0
2018 Q2	1.867967507	-27.15047482	-2.2565321	3.6	0
2018 Q3	1.963976467	81.50666265	-3.3214709	3.7	0
2018 Q4	2.573711669	587.1041255	-0.365408	3.3	0
2019 Q1	1.444456813	48.87057398	-2.3142509	3.6	0
2019 Q2	2.009553953	71.30950737	-1.7010936	3.6	1
2019 Q3	0.94431535	77.3813077	-1.595092	4	1
2019 Q4	-0.40907914	-136.9395595	3.3007335	3.7	1
2020 Q1	-2.937006396	449.3273624	5.7356608	1.6	1
2020 Q2	-25.54578441	238.2071121	18.665019	-14.5	1
2020 Q3	-11.06703083	-122.5340818	23.815461	-5.2	1
2020 Q4	-12.16333322	-116.3745685	17.278107	-4.9	1
2021 Q1	-12.61074452	-91.72388695	14.504717	-3.9	1
2021 Q2	24.86475432	-97.33737394	2.1875	18.2	0
2021 Q3	7.864753581	169.8506208	-3.6253776	8.1	0
2021 Q4	9.724414628	-207.7950851	-0.3027245	9.8	0
2022 Q1	14.68144139	0.164866348	0.2059732	12.6	0
2022 Q2	4.791453567	-0.038457981	-0.3058104	9.6	1
2022 Q3	1.46775269	0.142100435	2.0898642	8.4	1
2022 Q4	1.382886258	0.072842399	0.3036437	7	1

Source: Calculated based on secondary data obtained from the official website of Bank of England and Office For National Statistics, UK

1.6.1 Table 1 Interpretation

This table is the initial table on which we use logistic regression, and the following are its components:

- Four economics related data- Household final consumption, credit growth, public debt, and GDP (Gross Domestic Product) have been considered as the predictors(variables).
- Credit growth and public debt are taken as a ratio of GDP and then the change in the respective ratios over several quarters has been analysed. Quarterly growth of household final consumption and GDP is analysed on Year-on-Year basis.

1.6.2 Application of Logistic Regression

- Logistic regression is used to calculate the probability of a binary event (here, the occurrence or non-occurrence of a financial crisis) and deal with its classification.
- The data for Household Final Consumption, credit growth and public debt have been obtained on a monthly basis.
- They have then been converted into quarterly data represented by Q1,Q2,Q3 and Q4.
- In the following table, the meanings of the words are-
    - Intercept- it is the value where all the predictors (b1,b2,b3 and b4) are zero.
    - The logit function- this function represents the probability values from 0 to 1 and negative infinity.
    - e^{logit} is the exponent of the logit function, also known as "odds".
    - Probability is the likelihood of the happening of an event. It is calculated as-

$$\frac{e^{logit} \text{ or odds}}{(1+e^{logit} \text{ or odd})}$$

The cut off prediction is the cut off at which the logistic regression classifies the data into two simple outcomes, that is crisis or no occurrence of crisis, based on the probability calculated.

Table 2 – Application of Logistic Regression

LOGIT	e^{logit}- ODDS	PROBABILITY OF CORRECT MATCH	LOG LIKELIHOOD	CUT OFF PREDICTION	PROBABILITY OF OCCURANCE
-0.961447623	0.382339002	0.72341155	-0.323776994	0	0.27658845
-2.044919588	0.129390594	0.885433264	-0.12167819	0	0.114566736
-4.287649662	0.013737174	0.986448978	-0.013643675	0	0.013551022
-2.065073507	0.126808969	0.887461875	-0.119389717	0	0.112538125
-0.809804904	0.444944865	0.692067929	-0.368071165	0	0.307932071
-1.477179971	0.228280539	0.185853746	-1.682795227	0	0.185853746
3.58802785	36.16268729	0.973091289	-0.027277379	1	0.973091289
17.22489602	30246650.89	0.999999967	-3.30615E-08	1	0.999999967
23.30983255	13284085136	1	-7.5278E-11	1	1
73.09176352	5.53796E+31	1	0	1	1
65.52562399	2.86693E+28	1	0	1	1
56.93129023	5.30817E+24	1	0	1	1
56.03325393	2.16238E+24	1	0	1	1
-22.59781735	1.53424E-10	1	-1.53424E-10	0	1.53424E-10
-10.45556731	2.87876E-05	0.999971213	-2.87871E-05	0	2.87867E-05
-4.129207766	0.016095625	0.98415934	-0.015967464	0	0.01584066
-10.77916476	2.0829E-05	0.999979171	-2.08288E-05	0	2.08286E-05
11.53218909	101945.0365	0.999990191	-9.80916E-06	1	0.999990191
23.60352235	17818843330	1	-5.61203E-11	1	1
15.7410119	6858586.334	0.999999854	-1.45803E-07	1	0.999999854

Source: Calculated based on secondary data obtained from the official website of Bank of England and Office For National Statistics, UK

Table 3 – Intercepts and CoefficientInterpretation:

INTERCEPT	b1	b2	b3	b4
-2.914200892	-3.342723908	-0.0008392	2.023004	3.2377

The intercepts, and coefficients- b1, b2, b3 and b4 have been calculated and the regression equation has been optimised to the maximum extent by maximising the log likelihood as follows:

Table 4- Sum of log likelihood

SUM OF LOG LIKELIHOOD	-2.672659414

Another point to be noted is that the predicted probability of the crisis in the last three quarters of 2022 is very high, that is, around 99%. It reflects that using the quarterly data of the predictors, the logistic regression model could accurately estimate the likelihood of the crisis.

This further gives governments the opportunity to apply logistic regression on their forecasted values of the predictors to analyse how long the crisis is likely to continue, and therefore make necessary arrangements.

Table 5 – The probability of crisis occurrence after optimisation

2022 Q2	0.999990191
2022 Q3	1
2022 Q4	0.999999854

1.6.3 The Confusion Matrix

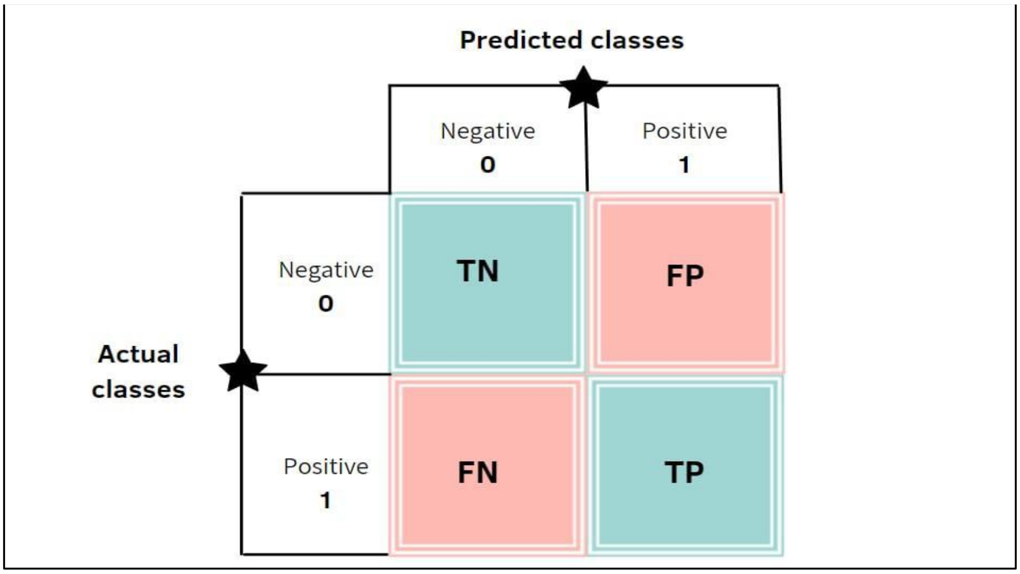
Table 6 – The confusion matrix of the data.
CONFUSION MATRIX

		Predicted 0	Predicted 1	
Actual	0	9	0	9
Actual	1	1	10	11
Total		10	10	20

The Confusion Matrix is a performance measurement tool in machine learning, that helps with further classification. It helps understand the probability of occurrences and matchers it with the actual occurrence of the event. The following is the Confusion Matrix generated by the logistic regression.

The Confusion Matrix can be interpreted as follows-

Table 7 – The confusion matrix interpretation



TP stands for True Positive and FP stands for False Positive at the cut off of 40%.

TN stands for True Negative and FN stands for False Negative.

- In the Table, we have –

○ 10 true positives- the prediction is true (that is the matrix predicted occurrence of financial crisis) and value is positive (that is financial crisis actually occurred).

○ 0 false positives- the prediction is incorrect, but the value is positive.

○ 1 false negative- the prediction is incorrect, (that is the matrix predicted occurrence of a crisis) and the value is negative (no crisis actually occurred).

9 true negatives- the prediction is correct (that is, the matrix predicted that no crisis would occur), and the value is negative (that is no crisis actually occured)**1.6.4 Accuracy, Precision and Recall**

Given is the table for accuracy, precision, recall, true positive and false positive:

Table 8– Accuracy, Precision, Recall, FP and TP

accuracy	95%
precision	100%
recall	91%
True Positive	91%
False Positive	0%

Interpretation

Precision - Precision refers to the ratio of truly positive outcomes over total positive outcomes. It attempts to answer the proportion of positive identification that were correctly identified out of total positive data set (that is the sum of true positives and false positives).

It can be calculated as-

$$Precision = \frac{TP}{TP + FP}$$

The precision value is 100% or 1. It indicates how precisely the model has predicted the occurrence of a financial crisis.

Recall - Recall is a metric that measures the percentage of actual positives that are accurately identified.

$$Recall = \frac{TP}{TP + FN}$$

The recall value is at 91%.

Accuracy - Accuracy shows how many times the prediction model was correct overall, by considering true positives and true negatives. It is given by-

$$Accuracy = (TP+TN)/(TP+TN+FP+FN)$$

The accuracy of the logistic regression model stands at 95%

1.6.5 Estimating the Right Cut off at Different Levels

Cut off is a classification method, where if the predicted probability (here, probability of occurrence) is greater than the specified cut off value, that observation is classified as a “positive” (or simply as 1).

This study decides the cut off value to be 40% as it gave the most optimised output of Accuracy, Precision and Recall. It is reflected in the table given below.

Table 9– Estimating Optimal Cut-Off Value

CUTOFF	ACCURACY	PRECISION	RECALL
0.00	0.55	0.55	1.00
0.10	0.80	0.73	1.00
0.20	0.85	0.83	0.91
0.30	0.90	0.91	0.91
0.40	0.95	1.00	0.91
0.50	0.95	1.00	0.91
0.60	0.95	1.00	0.91
0.70	0.95	1.00	0.91
0.80	0.95	1.00	0.91
0.90	0.95	1.00	0.91

1.7 CONCLUSION

The following conclusions can be drawn from this study:

- The model has been successful to predict the financial crisis using actual data of predictors.
- The probability of occurrence of financial crisis over the three quarters of 2022 has been predicted correctly.
- Economies can use logistic regression to understand the relationship between macroeconomic predictors and use it for policy making.
- The model has generated the highest possible rates of recall, precision and accuracy which are all above 90%. It means that the occurrence of financial crisis in such cases are high. This is also evident from the probability calculated for the Q1, Q2 and Q3 which is around 99%.
- Therefore, the same model can be applied on estimates of coming years to forecast how long the current crisis will continue and if used consistently in future years, it will help economies predict any impending financial crises.

References:

1. Kristina Bluwstein, Marcus Buckmann, Andreas Joseph, Sujit Kapadia, Özgür Şimşek (2021) – “Credit Growth, Yield Curve and Financial Crisis Prediction: Evidence from a Machine Learning Approach”. European Central Bank Working Paper Series, No. 2614/ November 2021.
2. Robin Greenwood, Samuel G. Hanson, Andrei Shleifer, Jakob Ahm Sørensen (2021) – “Predictable Financial Crises”, Harvard Business School Working Paper No. 20-130/ March 2021.