Aston Business School
BIRMINGHAM UK

INFLATION FORECASTING IN GREECE: A COMPARISON BETWEEN ARIMA, LSTM ARTIFICIAL NEURAL NETWORKS AND XGBOOST.

DIMITRIOS ANTONIOS KASTRINAKIS, 17 Jul 2023
MSc Business Analytics Student

Academic Supervisor: DR. FARBOD KHANIZADEH

Acknowledgements

This dissertation has been thoroughly guided by **Dr. Farbod Khanizadeh.**

Declaration

I declare that I have personally prepared this report and that it has not in whole or in part been submitted for any other degree or qualification. Nor has it appeared in whole or in part in any textbook, journal or any other document previously published or produced for any purpose. The work described here is my own, carried out personally unless otherwise stated. All sources of information, including quotations, are acknowledged by means of reference, both in the final reference section and at the point where they occur in the text.

Contents

Table of Contents

Abstract

Inflation has been plaguing societies since the ancient times, at occasions leading to complete market instability and significantly dropping the living standards of people in societies. In less severe scenarios, inflation adds uncertainty to investment decisions and shortens the investment horizon in emerging markets, making the generation of accurate inflation forecasting models imperative.

This paper seeks to compare three different widely used approaches in the domain of macroeconomic forecasting in determining which one of those approaches shows more promising results for short term inflation prediction. More specifically, the first method that is employed to forecast inflation is ARIMA, an extensively used, traditional, time series forecasting approach that focuses solely on past CPI values to generate predictions. Moreover, an LSTM Artificial Neural Network has been designed to forecast inflation, as a contemporary deep learning approach often used for regression tasks and time series forecasting. Finally, XGBoost, an ensemble algorithm that combines simple decision trees (base learners) is the third model that has been generated, which is the proposed method for inflation forecasting in this research project. All three methods are compared to a persistence model, which constitutes a challenging baseline to surpass in terms of its forecasting performance.

Findings show that, overall, XGBoost outperformed both the ARIMA and the LSTM ANN in terms of forecasting performance (RMSE). The ARIMA model demonstrated better performance than that of the LSTM, showcasing that simpler models may be more effective when working with only a limited amount of data. However, it should be highlighted that the LSTM model showed promising results in capturing unexpected inflation dynamics, particularly during extreme scenarios. On the other hand, although XGBoost exhibited the smallest RMSE score and most accurately predicted the direction of CPI changes, the model was completely unable to capture the extreme inflation spikes that occurred after 2022. The persistence baseline, although it exhibited the best results in terms of RMSE, it does not capture at all the direction of changes in CPI values, and hence its reliability as a forecasting model is compromised.

1. Introduction

1.1 Background

Machine learning is a branch of artificial intelligence that includes a large variety of methods that with only mild underlying assumptions about statistical relationships, assimilates past data in an effort to model those statistical relationships. In essence, machine learning uses a learning method and an algorithm that enable one to automate the generation of a predictive model, while the discretion of the forecaster is minimized (Hall, AS., 2018). Of course, it should be mentioned that it is quite challenging to model complex phenomena and select only the relevant variables, while at the same time choose an appropriate algorithm. On top of that, fine-tuning the hyperparameters also requires a lot of knowledge both on the background of the specific problem and on the algorithm's architecture. However, if one methodically generates an adequate, parsimonious, and good ML model, then the potential of this model can be so great that even the most complex economic phenomena are better understood and forecasted.

The objective of this project is to forecast Greek inflation based on selected macroeconomic and financial variables that are chosen according to relevant inflation theories. The goal is to examine whether XGBoost, a decision trees ensemble method, can improve the forecasting performance compared to an ARIMA model, a traditional macroeconomic forecasting method and a Long-Short Term Memory Artificial Neural Network (LSTM ANN).

Inflation is a phenomenon that has been plaguing economies since the ancient times. In ancient Rome for example there are records of rise in prices of basic goods due to increases in gold supply. For instance, during the 2nd and 3rd century, in the Region of Dalmatia, in only a century and a half, wheat price had gone up about 200 times, when gold supply in the same period had risen about 45 times (Jones A. H. M., 1953). Inflation spikes also appeared in the medieval times Europe (16th century) when vast quantities of gold and silver were imported to Europe from the Americas. That was also the first time that inflation was studied as a macroeconomic phenomenon by various scholars of the era (see section 2.1 – Inflation Theories). One rather exemplary instance of inflation, or perhaps hyperinflation, is that of post WW1 Germany, when the monthly inflation rate between 1922 and 1923 was an astonishing 322 percent. In essence, that meant that each month during this period, prices quadrupled (Backhouse, F., et al., 2023). Along with Germany, Greece was also one of the seven instances of hyperinflation that have occurred over the twentieth century. During WW2, the excessive reliance of the German puppet government on inflation tax led to an incredibly high inflation spike. Notably, when the stabilization of the Greek economy occurred, the Greek government's unit of account, the old drachma, was converted to the new drachma at a rate of 50 billion to one (Makinen, Gail E., 1986). Nowadays, inflation in Greece is much lower, at least in a historical perspective, however, empirical evidence shows that inflation adds uncertainty to investment decisions and shortens the investment horizon in emerging markets, making the generation of accurate inflation forecasting models imperative (Araujo, G. S., Wagner, P. G., 2020).

1.2 Definitions of inflation

If a non-economist person were to be asked what inflation is, it is quite possible that they would suggest that it is the increase in the overall prices. They could potentially add that products in the supermarkets are getting more expensive or that fuel prices are increasing. Others might also indicate that the wages are too low and hence, people's purchasing power has decreased. However, is any of the aforementioned laymen definitions, correct? Or if not, what is the universally accepted, scientific definition of inflation?
Unfortunately, there is no single way to define inflation and that is because of the complex nature of the concept itself. As a matter of fact, throughout time many Inflation theories have suggested different definitions, explanations and causes of inflation and hence, it is deemed useful to explore those theories in an endeavor to understand, define and perhaps predict inflation.

For example, one of the first definitions of inflation states that inflation is '*An increase in the amount of money issued beyond what is justified by the country's tangible resources.'.* (Collins dictionary, 1970). Thirty years later, Blanchard defined inflation as '*A sustained increase in the general price level of goods and services in an economy over a period of time*.' (Blanchard O., 2000). However, for the purpose of this research project, the American heritage dictionary's definition is deemed as the most appropriate, stating that inflation is *'A consistent increase in the level of consumer prices or a persistent decline in the purchasing power of money' (The American heritage dictionary, 2022).*

1.3 Forecasting inflation

It is well known in literature that attaining a good in-sample fit does not translate into a good out-of-sample performance (Greene, 2003). In other words, bias and variance usually go the other way and a mismatch between bias and variance can occur due to overfitting. In macroeconomics, a common hurdle is that measurements and indices are collected, in the best case, on a monthly basis, and hence, high-frequency datasets are scarce or in most of the cases, non-existent (Araujo, et al., 2020). It is reasonable then to infer that machine learning algorithms, and especially neural networks and decision trees, that are data-hungry, could in principle, have their predictive ability undermined. However, this is not the case only for this project but for most research papers regarding inflation forecasting methods, and hence, the performance of the model should be measured in the scope of the circumstances in the domain of macroeconomics. In any case, to ensure that the actual predictive power is tested, an out-of-sample empirical exercise with 4 inflation forecasting methods is designed. The forecasting horizon will be limited to 1 month, as the goal is not to predict inflation for a real-world application, but instead, to present the mere findings and conclusions by comparing different forecasting methods, that could potentially be of academic or even practical value. Two of the models that will be tested are considered traditional econometric approaches for forecasting inflation. That is a persistence model and an ARIMA model, the most widely used time-series model (Ang et al., 2007). Additionally, a deep learning model, more specifically, a development of the simple Recurrent Neural Network, a Long-Short Term Memory Artificial Neural Network (LSTM ANN) will also be employed. Finally, XGBoost, a widely used ensemble algorithm, is the proposed method that those models will be compared to.

1.4 Past Research

The literature on macroeconomic forecasting using machine learning methods is relatively new and in need of further development. Important pieces of research in the field of inflation forecasting, amongst others, include Araujo, G. S. & Wagner, P. G. (2020) paper, which constituted the foundation of this project, as it compared 16 different methods for inflation forecasting. Cheng et al. (2021), that aggregated survey-based forecast using ML tools to forecast US inflation, Nakamura E. (2005) that used ANN to forecast US inflation, and Paranhos L. (2021) that used LSTM ANN to forecast inflation.

This research may contribute to this fast-growing literature by (a) proposing that an XGBoost, which is a decision tree ensemble gradient boosting method may attain higher out-of-sample performance due to less overfitting in comparison to an LSTM ANN or other traditional macroeconomic methods, (b) be one of the few pieces of research that endeavor to forecast inflation via the use of ML methods in Greece and (c) variable selection is based on the suggestions of the leading macroeconomic theories in the field of inflation (quantity theory of money, Keynesian inflation theory and the inflationary wave theory), along with the use of other statistical methods for variable selection (i.e. Pearson's correlation, graphical approaches etc.).

2. Literature review

2.1 Inflation theories

2.1.1 The quantity theory of money

History of the theory

The cornerstones of the quantity theory of money were established as early as in the 16th century, after Christopher Columbus discovered the Americas. During that era money was defined in terms of precious metals such as gold and silver. With the gradual exploitation of the American land's resources by the Europeans, vast amounts of gold and silver were injected into the European markets, causing a shock to the European economy. As the amount of money circulated increased rapidly, prices of goods and services rose, and inflation spiked. Many philosophers and scholars of the era endeavored to understand the mechanics of the phenomenon and suggest any possible explanations. Amongst them, was the famous astronomer, Copernicus, who was one of the first to propose that price increases are related to increases in the money supply (Comley, P., 2015). Copernicus and his contemporaries, stated that Prices (P) are proportional to the Money Supply (M) and hence: $P \propto M$. This simple discovery became part of the very definition of inflation for a long time and set the foundations for the quantity theory of money.

Hume made a major contribution to political economy in his work ‘Political Discourses’ in 1752 (Hume, D., 1752; Schabas, M., et al., 2008). The famous economist was the first to suggest that an expansion of the money quantity may favorably affect an economy’s output and employment (Wennerlind, C., 2005). Although many critics argued that Hume’ s theory violated the neutrality of money condition (money is a ‘neutral’ factor that has no real effect on the economic equilibrium), a consensus interpretation gradually immerged suggesting that no such violation is in discussion and claiming that Hume’s monetary theory included a provision for the short-term non-neutrality of money(e.g. Vickers 1959; Stewart 1963; Blaug 1985; Hutchison 1988; Berdell 1996; Gatch 1996). Essentially, Hume’s work suggested that a small rate of an always increasing money supply may be beneficial for a society’s employment level and productivity. This is what we know today as ‘inflation-policy’. (Comley, P., 2015). For example, Switzerland’s, macroeconomic goal in the past years has been to keep inflation levels over 0 % and bellow 2% and has done so with a relative success (Capistrán et al., 2010).

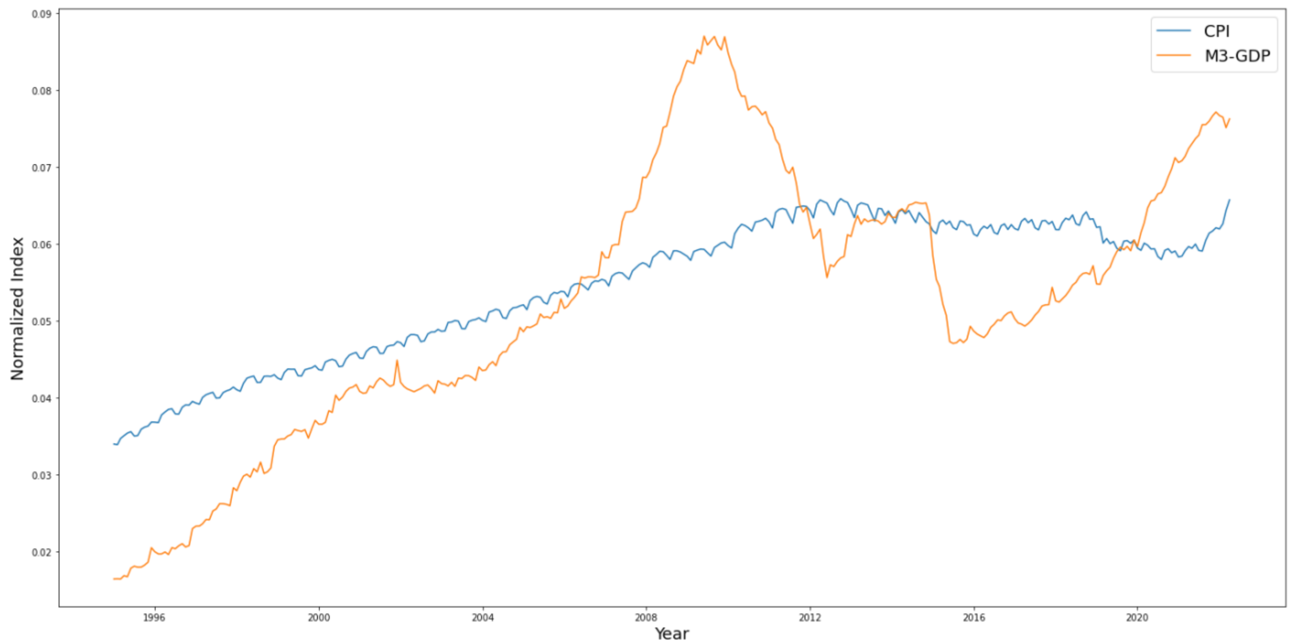
The equation of exchange

Perhaps the most significant development on Copernicus ideas is that of John Stewart Mill, when in 1848, the notable economist formally proposed the ‘Quantity theory of Money’ and the ‘Equation of Exchange’. Mill explained that the simple formula that Copernicus and his contemporaries suggested (**$P \propto M$**) would only be valid if the size of the economy is stable and at the same time there is no change in money velocity, meaning the number of times money is spent during a specific period by people. In other words, GDP should be constant and saving and spending levels should also be constant. Of course, this is impossible in a practical level and hence he proposed that **$MV = PQ$**, the equation of exchange (Mill, J.S., 1848; Dobija, M., 2011). It should not be omitted that although Mill conceived the equation of exchange, Irvin Fisher was the one to articulate its algebraic formulation as following (Fisher, I., 1912):

$$MV = PQ$$

M: Money Supply
V: Velocity of money
P: Average Prices
Q: Value of goods/GDP

The basic idea behind this formula is that at any time period, Money Supply (M), multiplied by the velocity of money (V), which is essentially the number of time money switches hands, is equal to the product of the total value of goods and services produced (GDP) and average prices. However, in practice, although money quantity theory might be applicable in the long run (in terms of decades), in the medium and short term the correlation seems to be verry little. As one can see in line-chart 1, M3(money supply) and CPI-GDP(GDP is deducted to comply with the equation of exchange), the two lines present little synchronisity especially in the short-term or they even go for many years in the opposite directions. In general, this might occur partly because factors such as people desire to save/spend(Velocity of money) is ommited and partly because the created money does not necessarily flow directly into goods and services. Overall, monetarists strict approach for inflation in terms of the equation of exchange might explain a lot about how inflation emerges, but cannot capture issues such as consumer sentiment and where increases in the money supply have gone to (i.e. asset prices, producer prices, retail prices etc). Additionally, although there might be a stronger link between inflation and money growth in high-inflation periods, this link seems to be rather weak in periods of low inflation. In a 2005 study that examined the correlation between money growth and inflation of all countries from 1969 until 1999, the conclusion was amongst others that high-inflation or hyperinflation countries in the sample showed a strong link between their inflation levels and money growth, while countries in the sample with low-inflation showed a weak or even absent relation between inflation and money supply (De Grauwe, P. and Polan, M., 2005).



Line chart 1-CPI VS Money stock. Designed in python, using ELSTAT data.

2.1.3 Asset Price inflation and interest rates

Asset prices

However, in essence, although fundamental, the quantity theory of money may be too simplistic to capture the complexity that is in the core of inflation. For example, a simple question that needs to be answered is ‘how is money supply expanded?’. Contrary to the common belief that money is simply ‘printed’ by the central banks, ‘…money is created by the banking system…’ (Bank of England Quarterly Bulletin, 2014), meaning private banks making loans to individuals or companies. However, the important thing to note here is that the vast majority of those loans are issued for purchases of houses, companies, bonds, stocks, shares, commodities and derivatives and only a very small proportion is granted for direct spending into the real economy (i.e., shopping, entertainment, travelling). This realization can be further reinforced by recognizing that less than 3% of UK money is cash (Jackson, A. & Dyson, B., 2013). In a paper published by the Bank of England in 2014, it is illustrated that money that is lent out is not first deposited by people. Instead, 97% of UK money has been literally created out of thin air by commercial banks when people or companies sought loans (Mcleay, M., Radia, A., Ryland, T., March 2014, Bank of England Quarterly Bulletin). It is important then to note that in a way, the quantity theory of money concerns two distinct economies: consumer spending and asset prices. However, money does flow between those 2 economies. For example, when one gets a loan to buy a house, the person that sells the house will eventually spend a significant proportion of the money they get in the real economy on goods and services. Money can also flow to the opposite direction. When people buy government bonds for example, money are transferred from the real economy to a savings account for a number of years. In any case, it is important to highlight that the duality of the economy's nature complicates and lags the relationship between money supply and inflation, and hence, indices pertaining to asset prices, national debt and house prices along with GDP, average prices and money stock should also be taken into account when trying to understand and predict inflation.

Inflation targeting

Inflation targeting's purpose is to control the economy, to make it more stable, and therefore more economically productive. The thinking is quite straightforward: In times of inflation, If a central bank increases the interest rates, private banks lend less money, as borrowing becomes more expensive, while at the same time, saving becomes more lucrative. Hence, money supply shrinks, thereafter demand shrinks and lower prices are attained. This can work the other way around too. If interest rates are lowered, borrowing becomes cheaper, banks lend more, people spend more money into the real economy and money supply increases and hence prices increase (Comley P., 2015). Therefore, it is eminent that interest rates might also be closely linked to inflation and could constitute a significant variable in inflation forecasting.

2.1.4 Keynesian Theory of inflation

Demand-Pull Inflation

Contrary to the strict 'simplistic' monetarist theory stating that money supply is a direct and proportional driver of inflation, John Maynard Keynes(1883-1946), in the early 20th century suggested that inflation is a rather complex phenomenon and money supply is only one of the various factors that can affect inflation levels in an economy. Keynes explains that when aggregate demand exceeds aggregate supply, inflation is pulled higher. With the term aggregate demand, Keynes includes consumption, investment and government expenditure and suggests that when the sum of those – especially in times of full employment and productivity – exceeds aggregate supply, prices rise. With that said, Keynes was indeed an advocate of fiscal policy, explaining that the government should intervene in the economy with higher taxes and less governmental spending, so that the volume of money in the economy is controlled and demand decreases. Keynes believed that in this manner the rise of inflation or even hyperinflation can be restrained (J.M. Keynes, 1936; Dmitrieva, O., Ushakov, D., 2011). Hence, Keynes brings forward factors such as governmental monetary policy and aggregate supply and demand as equally important in understanding, determining and hence forecasting inflation.

Cost-Push Inflation

However, Keynes did not stop there. He went another step forward, suggesting that inflation can be caused by the increasing costs of the private or public sector. He explains that labor unions may enforce wage increases and hence a more rapid increase in the wages than the increase in the productivity of labor is very likely to cause inflation. Keynes trail of thought is that wage increases lead to a higher cost of production and hence, in the case that productivity is stagnant, the prices of the products rise, causing an inflation domino effect (J. Totonchi, 2011). Keynes did not omit to explain that increase in the prices of imported raw materials and commodities may also cause a cost-push type of inflation. Finally, Keynes refers to a specific type of cost-push inflation, known as profit-push inflation or administered-price inflation, where oligopolist firms and monopolies (i.e., in the energy industry) raise their prices and hence inflation spikes (J.M Keynes, 1936 – From the collected Writings of John Maynard Keynes, 1971 book).

Inflation expectations & unemployment

Keynes and his contemporaries also suggested that inflation expectations can fundamentally drive inflation. Expectations of inflation, drive labor unions to enforce higher wages and businesses to raise their prices in an endeavor to forestall the expected inflation (Comley P., 2015). Hence, inflation expectations may result in an endless cycle of wage and price increases due to the fear of an increasing inflation and eventually lead to an always rising inflation level. Additionally, the Keynesian macroeconomic theory suggests an inversely proportional relationship between inflation and unemployment. This is very well illustrated in Philips Curve, that has been widely used in the past to depict the relationship between the change of wages and unemployment (Phillips, A., 1958; Stock, J.H, Watson, M.W., 2008), in order to forecast inflation (Arauju G., Gaglianone W. P., 2020).

2.1.5 New political Macroeconomics of inflation & interest rates

The major important theories as mentioned above mainly focus on the macroeconomic determinants of inflation and do not take into consideration societal and political non-economic factors that may constitute significant inflation drivers. Such factors include timing of elections, the level of political stability in a society, the reputation and credibility of policy makers, the performance of policy makers but also the inflation process itself. It is also important to add that the case for Central Bank independence is usually framed in terms of the inflation bias (deviation) present in the conduct of monetary policies. However, the theoretical and empirical work suggests that monetary constitutions should be designed to ensure a high degree of Central Bank autonomy. (C.A. Sims, 1980). In general, political, and social factors and the overall political climate that exists in a society, and of course the global political situation may be detrimental in inflation forecasting and unfortunately due to the inability to quantify those variables, it is still hard to accurately predict inflation.

2.1.6 Inflationary wave theory

A thorough examination of economic history suggests that there have been long periods, that may last even centuries that there is an overall price stability. On the other hand, there have been periods in history that inflation seems like a structural part of economies (Comley,

P., 2015). It appears to be a general secular trend towards a general increase in prices over the centuries driven by population increase and hence competition for resources (Malthus T. R., 1872; Lohmann, L., 2005). Pete Comley in his book Inflation Matters in 2015 explains in detail the inflationary wave theory that suggests that there is a marked wave pattern of a continually rising inflation for a long span of time (i.e., a century or more), followed by a period of stability in prices, before the cycle iterates.
Although there are not available official records of prices in Greece before 1959, for demonstration purposes, data of UK prices will be used that date back to 1264 ac. The chart bellow shows that the UK has experienced wave or price increases (shaded) followed by equilibrium periods of relative price stability. This pattern of course is not unique to the UK and it does occur with very similar timings across other European countries where records exist (Hackett Fishcer, D., 1999).

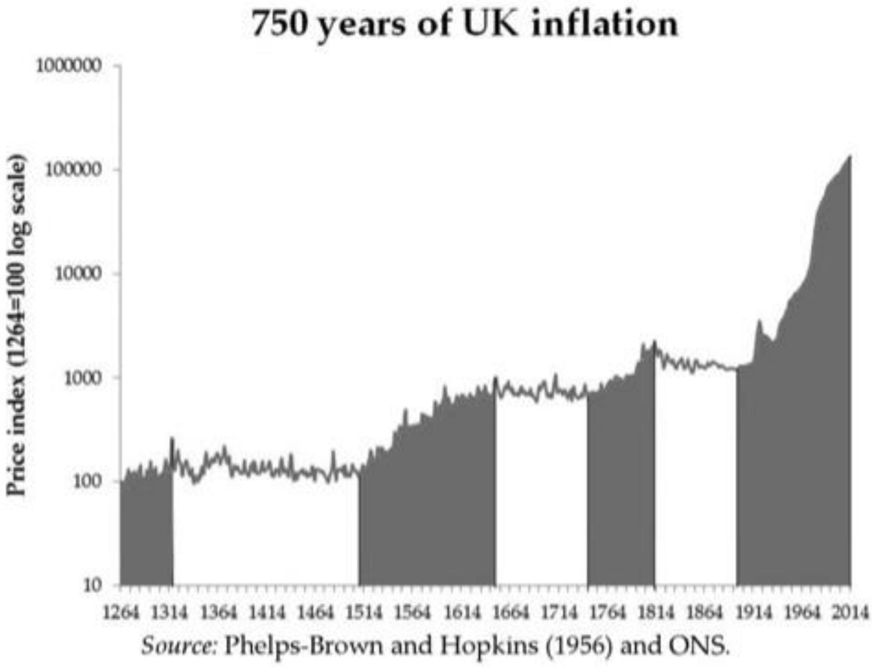


Figure 1 - Price index from 1264 to 2014 (source: Comley P., 2015).

The start and rise of the wave

One of the most detailed historical analyses of inflationary waves is that of David Hackett Fischer in 1999. Fischer hypothesized that after a period of price stability some factor triggers the beginning of a new inflationary wave (Fischer D., H., 1999). Fischer argues that the evidence supports the theory of T. R. Malthus in 1798, also known as Malthusianism. Malthus suggested that population at periods of ‘normality’ and ‘prosperity’ rises exponentially, whereas resources and commodities used for covering people’s basic needs only rise arithmetically. This mismatch usually leads to competition for those resources and hence prices rise (Malthus, T. R., 1872; Lohmann, L., 2005; Ross, E.B., 2003). Fischer goes on to argue that in times of overall stability, people perceive life more optimistically and consequently have more children. This gradually puts pressure on prices and the population growth itself leads to a gradual increase in prices over a long period of time. Fischer’s detailed analysis of economic history brings forward that it is not increases in the money stock that lead to a new trend, but instead, changes of the money supply usually follow the beginning of a new trend and amplify it. People realize that they live through an era of constantly increasing prices and an inflationary mind set takes in. Thus, people adopt behaviors such as demanding pay rises, the government expands the money supply to ensure there is enough money for the economy, and businesses increase their prices. In this manner the new inflation trend is reinforced.

The cycle’s termination and the price stability stage

The theory of "inflation waves" proposes that inflation tends to occur cyclically, much like waves in the sea. According to this theory, inflation builds up to a peak before collapsing, often in a catastrophic way, and is usually brought to an end by either war or population decline. Following this, there is typically a period of relative price stability, during which prices oscillate within a certain range and do not exceed the previous high. As time passes, the amplitude of these oscillations tends to decrease as a stable price mindset takes hold. During this phase, prices generally decline due to productivity and technological advances, which reduce the cost of production for certain goods. Assets such as land and houses also offer diminishing returns during this time, while the purchasing power of wages increases and rents decline. In the waves that follow a population decline, there is a scarcity of labor, which leads to higher wages as employers compete for workers from a smaller pool of candidates. This period results in a general decline in prices, with goods becoming cheaper and wages rising (Comley, P., 2015).

2.1.7 Theories & Time frames

Recapping all these theories leads to an interesting conclusion. Maybe all the theories are right to some extent. It is just the time frame of their influence that distinguishes them. Malthus's theory about demographics, population growth and competition for resources, along with Fischer's observations and Comley's ideas on inflationary wave theory have an impact over the very long-term inflation rate. Medium-term rates are influenced by money supply. Latent inflation created by printing money must be rectified at some point, but this might be delayed for decades. Short-term it seems to be factors expounded by Keynes that best explain the gyrations of price indices on a daily basis, e.g. commodity price rises and government-regulated price increases.

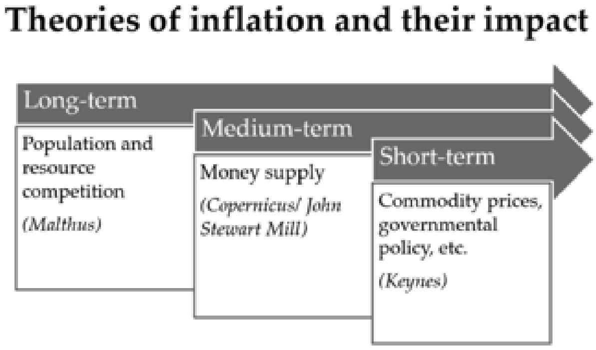


Figure 2 - Theories of inflation and their impact (source: Comley P., 2015)

2.2 Measuring inflation

(All the information bellow is based on Eurostat's HICP manual at ec.europa.eu)

There are two main inflation measures that are used in Greece:

- CPI (Consumer price index)
- HICP (Harmonized index of consumer prices)

The harmonized index of consumer prices HICP is produced by each EU member state according to specific regulations and a thorough methodology, outlined by the European Union. The aim of the index is to establish a context in which each EU member state can produce a high-quality measurement of consumer price inflation and compare it with other EU member states. The HICP is essentially a measure - or an index - of pure price change for goods and services. Simply put, it measures the change in the cost of a fixed basket of products over time. The term 'pure price change' means that only changes in prices between the current and the referenced period are measured by HICP. In terms of product coverage, HICP includes household final monetary consumption expenditures and non-consumption expenditures such as purchase of financial assets are not included into the scope of HICP. Additionally, by principle, certain expenditure categories are excluded such as that of imputed rentals for housing (the amount of rent one would pay if they did not own their residence).

Products included in the HICP	Products NOT included in the HICP
Food	Narcotics
Alcohol and tobacco	Imputed rentals for housing
Clothing	Game of chance
Housing	Prostitution
Household equipment	Life insurance
Health	Public insurance connected with health
Transport	Financial intermediation services indirectly measured
Communication	
Recreation and culture	
Education	
Hotels and restaurants	
Miscellaneous	

Table 1 - HICP divisions

These 12 divisions are then broken down into groups, classes and sub-classes.

Weights for the HICP sub-indices

The HICP sub-indices weights are essentially the proportion of aggregate expenditure on each set of goods and services included in the HICP categories over the total expenditure on all goods and services. Member states are required by EU regulations to update product weights for the HICP every year.

Index formulae

The HICP is an annually chain-linked Laspeyres-type index defined as:

$$P^{0,t} = \sum \frac{p^t}{p^0} \times w^{0,b}$$

Where,
P denotes the price of a product
p^0 denotes the price of the reference period
p^t denotes the comparison period
W (weights) denotes expenditure shares of a period b prior to the reference period and are adjusted to reflect the prices of the price reference period 0.

Differences between HICP and CPI

Although HICP and CPI do not differ much in their estimates, there are some fundamental differences in the methodology that is used to calculate each measurement. First of all, due to the harmonized methodology used to calculate the HICP, it allows for cross-country comparisons as opposed to CPI, which is set up to serve different national purposes. Hence, some of the underlying methods used to calculate CPIs may deviate from the requirements for the HICP. For example, Greek CPI does include imputed rentals for housing as opposed to HICP that does not. Additionally, HICP datasets are published by EUROSTAT whereas Greek CPI datasets are published by the Hellenic statistical company (ELSTAT).

2.3 Traditional Methods for inflation forecasting.

2.3.1 Overview

As articulated previously, inflation is a multifaceted phenomenon and different theories for over the past 300 years have tried to set a theoretical framework around inflation. From monetarists suggesting that money quantity is the main driver of inflation ((Mill, J., 1848; Irvin Fisher, 1911 etc.), to Keynesian macroeconomics suggesting that different factors such as supply and demand disequilibrium, wage inflation or administered-price inflation can lead to price increases (e.g. J. M. Keynes, 1936; J. Totonchi, 2011). From modern theories suggesting that political & social developments affect inflation or that inflation follows a wave pattern, and finally to other theories arguing that inflation expectations and past inflation prices affect the future inflation values. It is eminent that different analysts and experts endeavor to forecast inflation based on different methods and models suggested by different theoretical frameworks. There is a variety of approaches in the literature to model the inflation dynamics. According to Ang et al. (2007), economists use four main methods to forecast inflation: time-series models(looking at past inflation values), structural models (e.g., Phillips curve), asset price models (e.g., term-structure of interest rates), and methods that employ survey-based measures (e.g., survey of professional forecasters). However, with recent developments in machine learning, analysts and economists endeavor to forecast inflation using various advanced non-linear ML models such as Random Forest, Recurrent Neural Networks, Decision trees and others.

2.3.2 Persistence Baseline

The persistence model does not constitute an accurate way in predicting inflation, at least in terms of directional accuracy but it is a good baseline model to compare the results of other models to. Nevertheless, due to the fact that inflation is considered a rather persistent measurement, a persistence model, that essentially shifts past inflation values one step ahead and uses the shifted values as the forecasted inflation, may actually lead to a lower RMSE score compared to other approaches for inflation forecasting. However, it should be highlighted that when predicting inflation, the direction of changes is equally important to the magnitude of the inflation measurement and hence, low RMSE does not automatically translate to a good and parsimonious inflation model (Xia, F., et al., 2019).

2.3.3 ARIMA

Unlike other models, ARIMA methods for time-series forecasting are considered agnostic. In essence, they do not assume some sort of underlying knowledge regarding any economic model or relationships and consider only past values of the time series in combination with previous error terms to make forecasts. Only requiring data from the time series in question is one of the greatest advantages of ARIMA, especially if one is forecasting lengthy time series. Additionally, this avoids the problem that usually occurs with multivariate time series, where consistent time series between different variables only occur for a shorter period of time. For example, in the case of inflation, CPI data date back to 1959, while there are only available M3 data since the 1980s. If a multivariate model was to be used, the data points would shrink dramatically compared to an ARIMA model. Finally, timeliness of data can be a problem for multivariate structural models. More specifically, if some of the variables included in a large structural model are published with a long lag, let's say wage data, then forecasts based on forecasts of unavailable observations may add additional uncertainty to the model (Meyler et al., 1998).

On the other hand, ARIMA models come with their downside. For example, in comparison with other traditional model identification techniques, ARIMA models are not based on the subjectivity of the forecaster and hence, their skill and experience cannot affect to the same extent the results of the model. On top of that, the economic significance of an ARIMA model is not clear as it is not embedded within any structural relationship or theoretical econometric model. Finally, because ARIMA models are essentially backward looking, they are generally showing poor results in forecasting turning points (Frain, J., 1992).

Overall, ARIMA models have proven themselves to be quite robust and efficient in short-term inflation forecasting. In many cases, ARIMA models may outperform more sophisticated multivariate traditional models in terms of short-term inflation forecasting (i.e., see Stockton and Glassman, 1987; Litterman, 1986). In other cases, however, previous research shows that more advanced ML algorithms may outperform ARIMA in terms of RMSE scores in inflation forecasting for all time frames. Nevertheless, it also looks like ARIMA models constitute a great benchmark for inflation forecasting evaluation (see Araujo, G. S., Galianone, W. P., 2020; also see appendix 1).

Although the process of estimating a general ARIMA model is not sequential and can involve iterative loops depending on the results of the diagnostic and forecasting stages, in this part a general outline of ARIMA modelling and forecasting strategy will be illustrated. First and foremost, data collection and statistical examination takes place. Thereafter, the stationarity of the data is to be determined. As soon as the data are differenced and rendered stationary, the correct ARIMA model must be identified and estimated. This can usually be done with the BOX-Jenkins methodology. Thereafter, the estimated model should be run with diagnostic checking (usually checking the residuals) and undergo sensitivity analysis. Should any problems arise, one should return to the model identification stage and re-estimate the model. Once the model has been chosen, then it can be used to forecast the time series, ideally using out-of-sample data to evaluate the forecasting performance of the model (Central Bank and Financial Services Authority of Ireland, 1998). The figure bellow illustrates the aforementioned procedure. (See Meyler et al., 1998, for a more in depth illustration of the process).

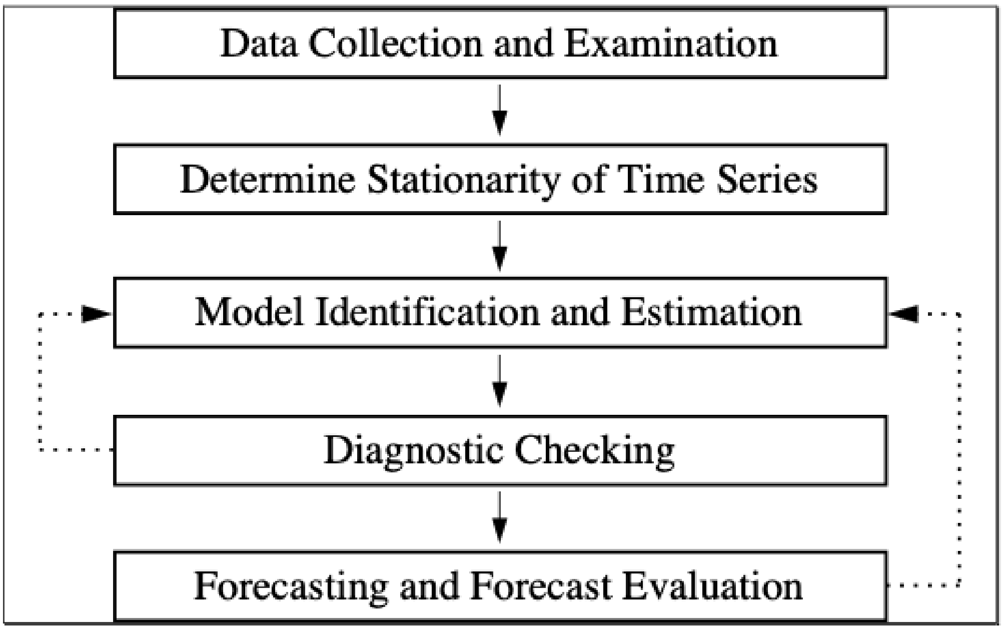


Figure 3 - ARIMA forecasting procedure (source: Wenying, J., 2018)

2.4 Artificial Neural Networks for inflation forecasting.

2.4.1 Overview

Neural Networks have been used in the context of multiple applications, aiming at modelling the dynamics of rather complex systems. Even though there is a wide range of Neural network types, their modelling accuracy is directly linked and dependent on the fit of the ANN architecture and considered problem. The aim of this report being to forecast future inflation values in Greece, it is suggested that a specific type of Recurrent Neural Networks, a Long-Short-Term-Memory Neural Network (LSTM) will be used in making the forecasts due to the nature of its architecture. It has been indicated that LSTM is a good candidate for setting up prediction models based on time-series data or data sequences to predict nonlinear time-variant system outputs (Lindermann, B., et al., 2020).

2.4.2 Recurrent Neural Networks

Although Recurrent Neural Networks have been proven efficient in various sequence learning tasks (Werbos, 1988; Schmidhuber, 2015; Rumelhart, et al., 1985), the difficulty in training such models has brought forward the need for improving their core architecture (Bengio, et al., 1994; Pascanu, et al., 2012). One of the most successful variants of RNN is that of Long short-term memory Neural Networks (Hochreiter, Schmidhuber, 1997). The LSTM architecture allows the network to store and retrieve information over long time periods with the use of gating mechanisms.

Just like other neural network architectures, recurrent neural networks have weights, biases, layers, and activation functions. The big difference is that recurrent neural networks also include feedback loops. Feedback loops allow the use of sequential input values, like past CPI values, to make inflation predictions. Essentially, the simplest instantiation of a recurrent neural network works as following: Let be the model's input time series consisting of T data points and be the model's results also consisting of T data points. A basic RNN unit would be defined by the following equation:

$$r_t = \tan\mathrm{h}(x_t u_t + r_{t-1} w + b),$$

Where u, w and b are the parameters of the model and tanh(x) is the model's activation function. The type of activation functions usually used in vanilla recurrent neural networks (although not limited to) is the hyperbolic tangent functioned defined as:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

It is important to note that r_{t-1} is used as an input as well as x_t. This is key to the model's efficiency in sequence learning tasks (Barkan, Oren, et al., 2022).

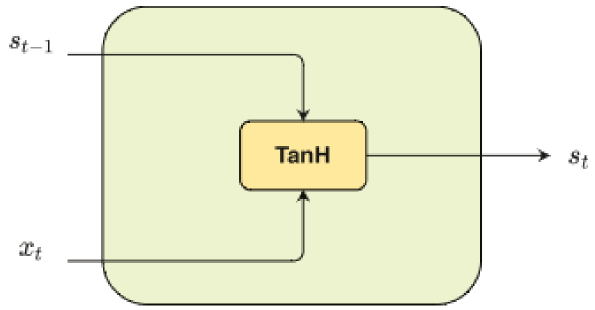


Figure 4 - Simple RNN unit (Source: Barkan, O., et al., 2023)

2.4.3 Long short-term memory Neural Networks (LSTM)

One common and serious problem that recurrent neural networks may encounter is the vanishing/exploding gradient (see Hochreiter et al., 2001). In order to deal with this problem amongst other optimization hurdles, LSTM Neural Networks have emerged, as they are tailored to overcome such hurdles. Due to LSTM Neural Networks effectiveness in dealing with such optimization hurdles, such models have found many applications such as: handwriting recognition (A. Graves et al., 2009; V. Pham et al., 2013; P. Doetsch et al., 2014) and generation (A. Graves, 2013), language modeling (Zaremba W. et al., 2014) and translation (M.-T. Luong, I. Sutskever, Q. V. Le, O. Vinyals, and W. Zaremba.,2014), acoustic modeling of speech (H. Sak, A. Senior, and F. Beaufays, 2014), speech synthesis (Y. Fan, Y. Qian, F. Xie, and F. K. Soon, 2014), protein secondary structure prediction (S. K.

Sønderby and O. Winther, 2014), analysis of audio (E. Marchi, G. Ferroni, F. Eyben, L. Gabrielli, S. Squartini, and B. Schuller, 2014) and video data (J. Donahue et al., 2014) among others.

The LSTM initial version included cells, input and output units but it did not include a forget gate or peephole connections. Training was done by including a mixture of back propagation through time and real-time recurrent learning (Robinson, A. J., Fallside, F., 1987; P. J, Werbos, 1988; Williams, R. J., 1989). The first paper to suggest a modification of the Vanilla LSTM architecture, included a forget gate that enabled LSTM to reset its own state before processing a new input (Gers, F. A., Schmidhuber, J., Cummins, F., 1999). Followingly, Gers and Schmidhuber (Gers, F. A., Schmidhuber, J., Cummins, F., 2000) also argued that so that a model can learn precise timings, the cell needs to control the gates and thus peephole connections were added to the architecture. Finally, in 2005 Graves and Schmidhuber (Graves, F. A., Schmidhuber, J., Cummins, F., 2005), presenting the full back propagation through time (BPTT) training for LSTM networks and presented results on the TIMIT (J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, 1993).

A schematic of the vanilla LSTM block can be seen bellow in fig. 4. It includes an input, output and forget gate, along with a block input and a single cell, peephole connections and an output activation function. It is eminent that the output block is recurrently connected to all of the gates and the input block. Let X^t be the input vector at time t, N be the number of LSTM blocks and M the number of input. The inputs weights are: $\mathbf{W_z}, \mathbf{W_s}, \mathbf{W_f}, \mathbf{W_0}$.
The recurrent weights are: $\mathbf{R_z}, \mathbf{R_s}, \mathbf{R_f}, \mathbf{R_0}$. The peephole weights are $\mathbf{p_z}, \mathbf{p_s}, \mathbf{p_f}, \mathbf{p_0}$. The bias weights are $\mathbf{b_z}, \mathbf{b_s}, \mathbf{b_f}, \mathbf{b_0}$. Hence, the vector formulas for the LSTM layer forward pass can be written as:

$$\bar{\boldsymbol{z}}^{t} = \boldsymbol{W_z x^t} + \boldsymbol{R_z y^{t-1}} + \boldsymbol{b_z}$$

$$\bar{\boldsymbol{z}}^{t} = \boldsymbol{g}(\bar{\boldsymbol{z}}^{t}) \qquad \textbf{block input}$$

$$\bar{\boldsymbol{\iota}}^{t} = \boldsymbol{W_i x^t} + \boldsymbol{R_i y^{t-1}} + \boldsymbol{pi} \odot \boldsymbol{c^{t-1}} + \boldsymbol{b_i}$$

$$\bar{\boldsymbol{\iota}}^{t} = \boldsymbol{\sigma}(\bar{\boldsymbol{\iota}}^{t}) \qquad \textbf{input gate}$$

$$\bar{\boldsymbol{f}}^{t} = \boldsymbol{W_f x^t} + \boldsymbol{R_f y^{t-1}} + \boldsymbol{p_f} \odot \boldsymbol{c^{t-1}} + \boldsymbol{b_f}$$

$$\bar{\boldsymbol{f}}^{t} = \boldsymbol{\sigma}(\bar{\boldsymbol{f}}^{t}) \qquad \textbf{forget gate}$$

$$\boldsymbol{c^t = z^t \odot \bar{\imath}^t + c^{t-1} \odot f^t} \qquad \textbf{cell}$$

$$\boldsymbol{\bar{o}^t = W_o x^t + R_o y^{t-1} + p_o \odot c^t + b_o}$$

$$\boldsymbol{\bar{o}^t = \sigma(\bar{o}^t)} \qquad \textbf{output gate}$$

$$\boldsymbol{y^t = h(c^t) \odot o^t} \qquad \textbf{block output}$$

Where σ, g, h are non-linear activation functions. Pointwise multiplication of two vectors is denoted by ⊙ **(**Klaus Greff, Rupesh K. Srivastava, Jan Koutn ́ık, Bas R. Steunebrink, Jürgen Schmidhuber, 2017**).**

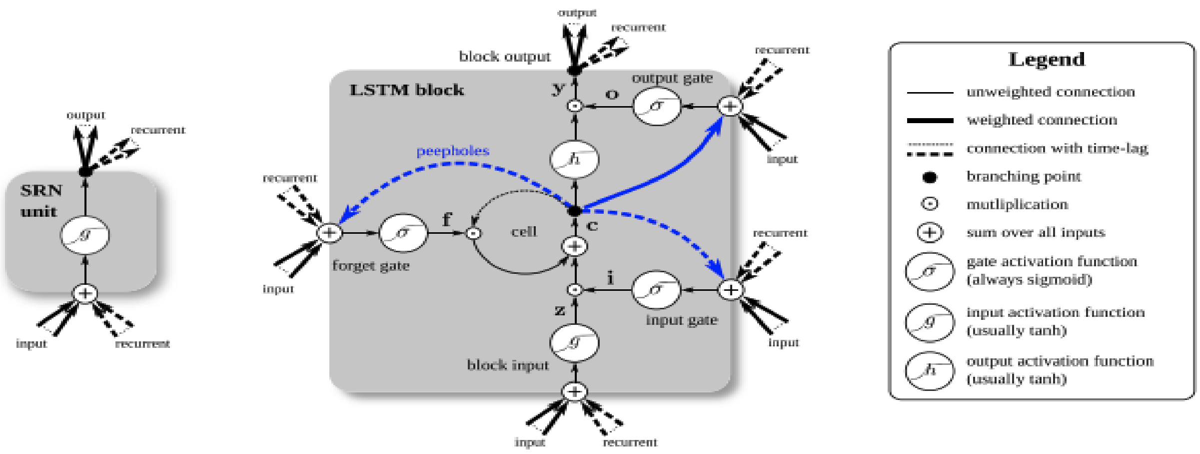


Figure 5 – Detailed schematic of the SRN unit (left) and an LSTM block (right) as used in the hidden layers of a recurrent neural network. (source: Greff, K., et al., 2017).

2.5 XGBoost regression

2.5.1 Overview

In general, ensemble learning combines individual models during training, namely base learners, to get a single prediction. XGBoost (extreme gradient boosting) belongs to the family of gradient boosting algorithms and is an ensemble learning approach, used to train many ‘weak’ models, usually decision trees and can produce a single best output (Tianqi C., Guestrin C., 2016). Boosting algorithms are essentially an ensemble method that may constitute a useful tool that can be applied in different predictive tasks and perhaps consistently provide higher accuracy results compared to single strong ML models. In the case of inflation forecasting for example, it is suspected that due to the high flexibility and customizability of decision trees ensemble boosting methods such as XGBoost, the easy application of such algorithms and the consistently high performance in comparison to other methods, XGBoost may be a good approach to model inflation (Bissacco et al., 2007; Hutchinson et al., 2011; Pittman and Brown, 2011; Johnson and Zhang, 2012).

2.5.2 Gradient boosting – an ensemble method

The most common approach to data-driven problems is to build a single, good, and parsimonious predictive model that captures the relationship amongst the variables and is able to produce a future prediction. In contrast to that, the main idea of boosting is to sequentially add new models and create an ensemble. In essence, at each iteration, a new, weak, base-learner model is trained to correct the error of the whole ensemble so far and improve it (Freund and Schapire, 1997; Friedman et al., 2000; Friedman, 2001).

In gradient boosting machines (GBMs), the principal idea behind this type of algorithms is to construct the new base-learners (the decision trees) based on the correlation with the negative gradient of the loss function, associated with the whole ensemble (Natekin A., Knoll A., 2013), as it points towards the direction of improvement. In the case of this project, the loss function is set to be the classic root mean squared error loss function.

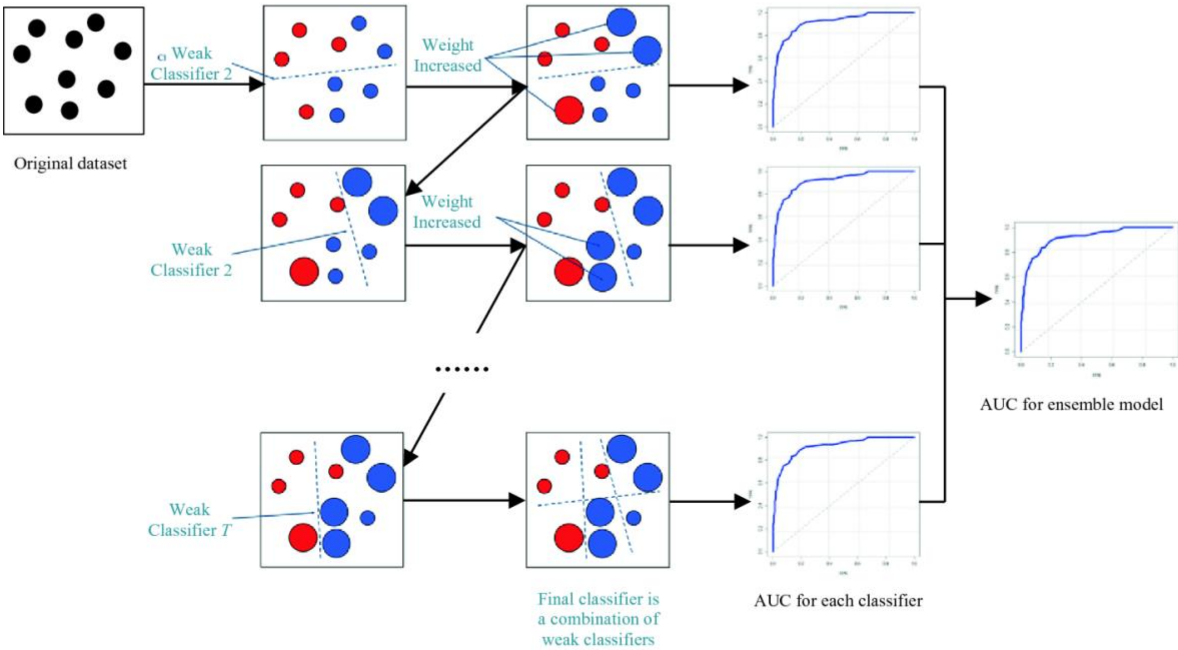


Figure 6 - Gradient boosting process (source : datascience.eu)

2.5.3 Extreme Gradient boosting – XGBoost

As previously explained, Gradient boosting is a general ML term referring to the technique of sequentially combining different weak models, typically decision trees, to create an ensemble of those predictive models and potentially, make better predictions. Extreme Gradient boosting is a specific implementation of this technique that takes use of several enhancements to improve performance and scalability. For the rest of this section, some of XGBoost's basic characteristics will be analyzed, as explained by Chen and Guestrin, in their paper, XGBoost: A scalable Tree Bosting System, in 2016.

Approximate Greedy Algorithm

Most GBMs including the simple machine version of XGBoost take use of the Basic Exact Greedy algorithm in determining the best split in tree learning. In essence, this algorithm iterates over all the possible splits on all the features. In order to do so, the Greedy algorithm sorts the data according to feature values and makes consecutive splits to accumulate the gradient statistics in order to evaluate the split candidates.

In XGBoost the algorithm used is slightly different. Although the exact greedy algorithm is extremely powerful as it looks through all possible splitting points, at the same time is computationally inefficient. Hence, XGBoost utilizes the Approximate Greedy Algorithm that first proposes candidate splitting points according to quantiles of feature distributions and then maps the features into bucket splits based on these candidate points, in order to aggregate the statistics and reach the best solution among proposals based on the aggregated statistics. By default, the approximate greedy algorithm uses about 33 quantiles.

Weighted Quantile Sketch

Now the question is, why is the phrase 'about 33 quantiles' used instead of 'exactly 33 quantiles' ? In answering this question, the concepts of parallel learning and Weighted Quantile sketch need to be investigated. When the amount of data is too large to fit the memory, simple calculations like finding the quantiles becomes extremely slow and hence a class of algorithms called Sketches can quickly create approximate solutions (see Jiang, Jiawei, et al., 2018 for more details) using parallel learning. In essence, the input data are split into multiple subsets (sketches) that are independently processes in parallel. Each sketch only captures an approximation of the data, allowing efficient computations. Then those sketches can be combined to obtain an approximation of the desired computation in the dataset and hence the quantiles are approximately 33. It is important to highlight that the quantiles used are weighted quantiles, as Chen and Guestrin explain in their paper, XGBoost: A scalable Tree Bosting System.

Sparsity-Aware Split Finding

In most real-world problems, it is common that some features' values are sparse. This may occur due to the presence of missing values, due to future engineering outcomes or any other possible reasons. XGBoost, as Chen and Guestrin explain, handles this issue by making the algorithm 'aware' of the sparsity pattern in the data. More specifically, when an input is missing, the instance is classified into the default direction, meaning the optimal direction learnt from the data by calculating the similarity score of each decision tree in different scenarios. It is important to note that sparsity aware algorithm is much faster and computationally efficient than the naïve version, in sparse datasets, as Chen and Geustrin explain.

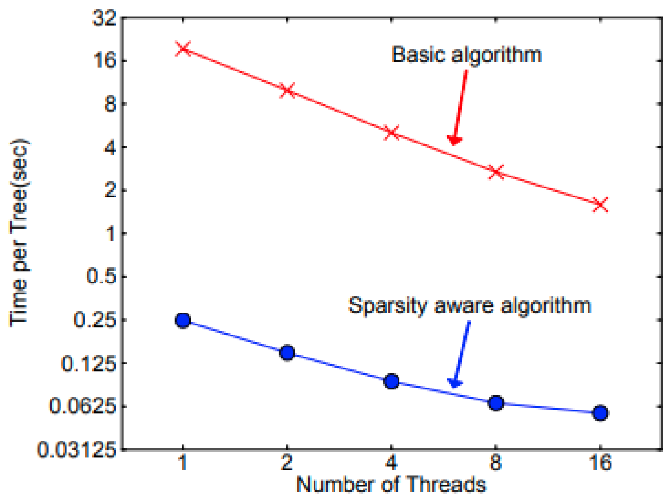


Figure 7 - Impact of the sparsity aware algorithm on Allstate-10K. The dataset is sparse mainly due to one-hot encoding. The sparsity aware algorithm is more than 50 times faster than the naive version that does not take sparsity into consideration. (source : datascience.eu)

Cache-Aware Access and blocks for Out-Of-Core Computation

It shouldn't be omitted that XGBoost instead of using the computer's main memory or hard drive, stores the Gradients and Hessians in the Cache memory, so that it can rapidly calculate similarity scores and output values. This advancement has enabled XGBoost to be even more computationally efficient, along with the Approximate Greedy algorithm, parallel learning, quantile sketch and sparsity aware algorithm.

Additionally, when the dataset is too large for the Cache and Main memory, then, at least some of it, must be stored on the Hard drive. However, because the Hard Drive is slow to read and write data, XGBoost minimizes these actions by compressing the data before reading and writing them, a process that is faster than reading and writing the actual data. Moreover, when there is more than one Hard Drives available, XGBoost uses a database technique called Sharding to speed up disk access.

3. Methodology and Analysis

The theoretical framework that will be used for this project is CRISP-DM. CRISP-DM is a process brought forward as an EU information technology project that is now the most widely used data mining model (SV-Europe, 2019). The model guides are available online as they are published by IBM. The figure bellow presents the six sequential stages of CRISP-DM (Chapman, 2019):

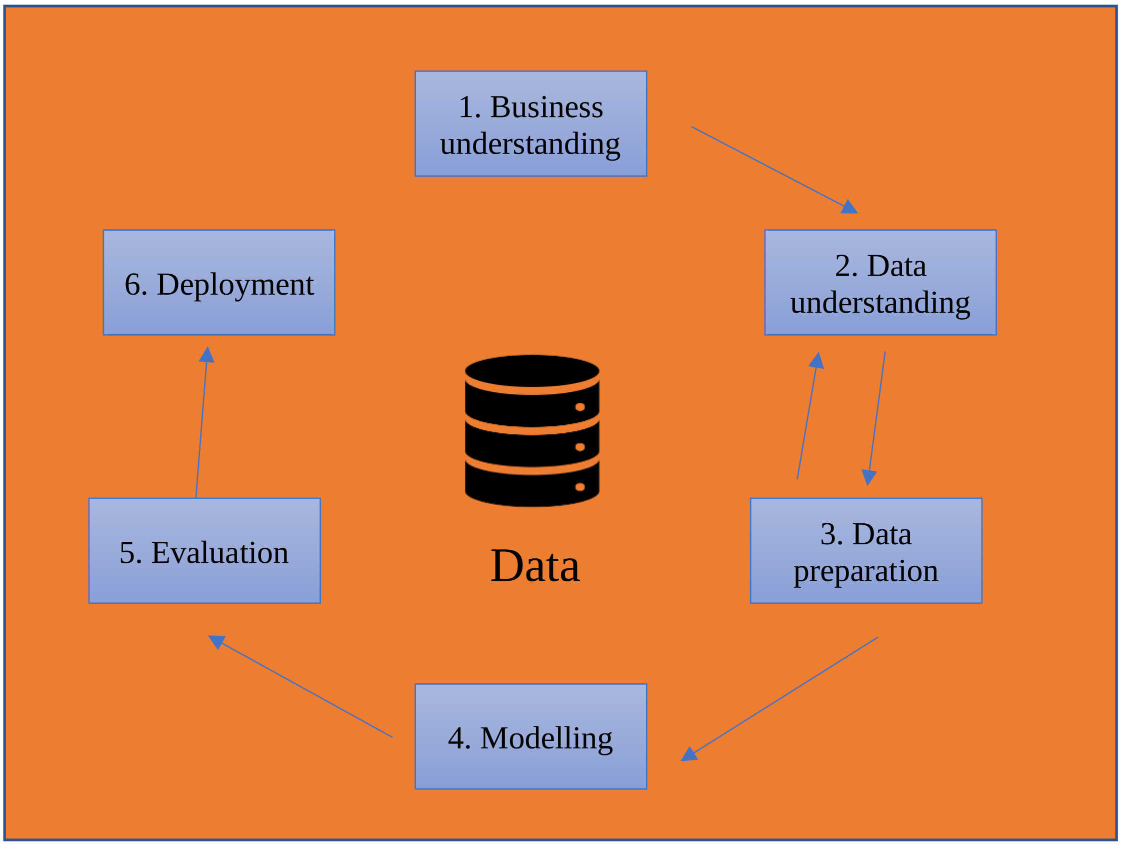


Figure 8 - CRISP-DM process

3.1 Business Understanding

In the pursuit of generating a complete, parsimonious, adequate, and optimal model that will be able to forecast future inflation values in Greece, it is important to understand the objective of macroeconomists that study inflation and the reasons why inflation is studied and forecasted, along with who will be benefited by an unbiased forecast of future inflation values.

Instances of potential beneficiaries from this project are illustrated bellow:

Private sector:

It is eminent that in times of uncertainty, businesses may take less informed decisions. Uncertainty in businesses may be caused by various internal or external factors. Inflation is an external factor that reinforces the uncertainty for most businesses in most industries. From asset vulnerability to price rises in commodities, raw materials or even basic goods, inflation as explained in section 2 of this report can majorly influence most aspects of the economy and hence companies in most industries must be ready to expect inflation and forestall it, in an endeavor to decrease their loss. Of course, this must be done with caution, since inflation expectations tend to rise inflation (see section 2, literature review).

Public policy:

As previously explained, public policy and central bank decisions can influence inflation levels drastically. Governmental spending levels and interest rates are two of the main weapons that are used to inhibit inflation. However, it is important that they are used in advance of inflationary waves and at the right level, otherwise the results may lead to catastrophic events in the economy. Raising interest rates too much for example, may lead to the bankruptcy of banks, companies, or even whole countries. Being able to forecast future inflation values accurately, can enable central governments and banks to make more informed decisions.

People:

It is eminent that the effects that inflation has on businesses and governments and the decisions that the private and public sectors take to forestall inflation affect the citizens of the Greek economy immediately. Therefore, accurate inflation forecasts and adequate situation handling is detrimental in sustaining a healthy and developed national economy.

The objectives of this report, in an endeavor to predict future inflation levels for the benefit of citizens, companies and the government are the following:

1. Understand the different theoretical frameworks that explain inflation (section 2 - literature review).
2. Discover the variables that may be linked or cause inflation, according to inflation theories (section 2 - literature review) and other statistical methods (section 3.2 – Data understanding).
3. Visually present the data of the selected variables in an endeavor to better understand relations (section 3.2 – Data understanding).
4. Clean and verify the datasets selected for the analysis (section 3.3 – Data Preparation).
5. Generate a parsimonious, adequate, and robust LSTM Artificial neural networks that forecasts future CPI values (section 3.4.5).
6. Generate a parsimonious, adequate, and robust XGBoost model that forecasts future CPI values (section 3.4.6)
7. Compare the generated models with traditional methods (ARIMA and Persistence model) to assess whether the model contributes to inflation forecasting or has the capacity to contribute with future improvements.
8. Reach an informed conclusion and make future recommendations based on the results of the analysis and the comparisons made.

3.2 Data Understanding

“Collecting data from data sources, exploring, and describing it and checking the data quality are essential tasks in this phase. To make it more concrete, the user guide (IBM’s CRISP-DM user guide) describes the data description task with using statistical analysis and determining attributes and their collations.” (Christoph Shröer et al., 2020).

According to CRISP-DM framework, data exploration and visualization is detrimental in understanding what the data is all about and what is noteworthy and interesting about the data (Knaflic, 2015). For this project's exploratory analysis, univariate and multivariate visualizations are generated using Python and all notable graphs are presented. Additionally, measures of dispersion and central tendency are also examined and potential correlation tests between different variables are generated (using Pearson’s correlation coefficient). In this manner, the modelling is done with much more confidence, as apparent relationships amongst predictors and between predictors and the target variable are noted during the EDA. In general, knowing the quality, quantity and relations of a dataset is a detrimental and integral step towards preparing the data and using them for model generation. Data visualization and EDA will also be useful in detecting outliers or missing values and find relevant treatments.

3.2.1 Collecting initial Data.

The data that have been collected are existing secondary data, retrieved from the Greek statistical authority (ELSTAT), from the European statistical authority (EUROSTAT), but also from the Greek and European Central Banks and the World Bank. Table 2 presents all predictor categories along with their specific predictors and where the data has been retrieved from.

Table 2 - Potential inflation predictors

Category	Potential Predictors
Inflation	**CPI**, HICP (Collected from ELSTAT and EUROSTAT)
Banking sector	Total Deposits, loans to private sector (Collected from the Central Bank of Greece)
Economic activity	Quarterly GDP, capital flows (Collected from ELSTAT)
Commodity prices	Commodity price data(69 different variables), electricity consumption (Collected from the World Bank and ELSTAT)
Exterior	Import price index. (Collected from ELSTAT)
Financial sector	Greek bonds maturity, Athen’s stock market general index, house price index (Collected from ELSTAT)
Industry and agriculture	Business confidence, capacity utilization, industrial production, PPI (Collected from ELSTAT)
Interest rates	Nominal interest rates, market interest rates (Collected from ELSTAT)
Labor	Unemployment, wage index (Collected from ELSTAT)
Money	M1, M2, M3, EU monetary base (Collected from ELSTAT and EUROSTAT)
Turnover indices & National economy	Retail turnover index (Collected from ELSTAT)
Inflation expectations	OECD Inflation expectations (Collected from OECD)
TOTAL # OF POTENTIAL PREDICTORS	**26**

3.2.2 Storytelling with Data

According to CRISP-DM user guide, after the initial data has been collected, the next step is to focus on the quantity and quality of the data at hand. Questions such as, what is the condition of the data, how much data is available and what are the value types or coding schemes are critical.

More specifically, the Amount of data, along with the frequency of each dataset is very important in time-series forecasting especially when machine learning methods are employed. Therefore, it is important that size statistics, the number of records and the number of fields, along with the frequency of data are noted. Additionally, not paying attention to the value types of the data (numeric, categorical, or Boolean), can lead to problems in the modelling phase of the experiment. Finally, it is imperative that attention is also drawn in this phase of the report on the various coding schemes that are potentially used in the datasets. For example, in one dataset Male and Female may be denoted as M or F and in another as 1 and 2. Any conflicting schemes should be noted and then be treated in the data cleaning process. Additionally, it is vital that basic statistics such as mean, median, mode, standard deviation and count are used in assessing the quantity and quality of the data.

Overall, generating descriptive statistics and univariate and multivariate visualizations, will bring forward the answers to integral questions that need to be asked before proceeding to the modelling phase of this project.

Such questions include:
Which ones of the variables seem to be of high quality and quantity, and can potentially be used for inflation forecasting? Which ones of the variables seem to be related in any way to the target variable? What kind of hypotheses have been formed about the data? Which attributes seem more promising for further analysis? Are there new characteristics about the data revealed? Do the data seem to come into agreement with the literature, or do some of the initial hypotheses need to be changed?

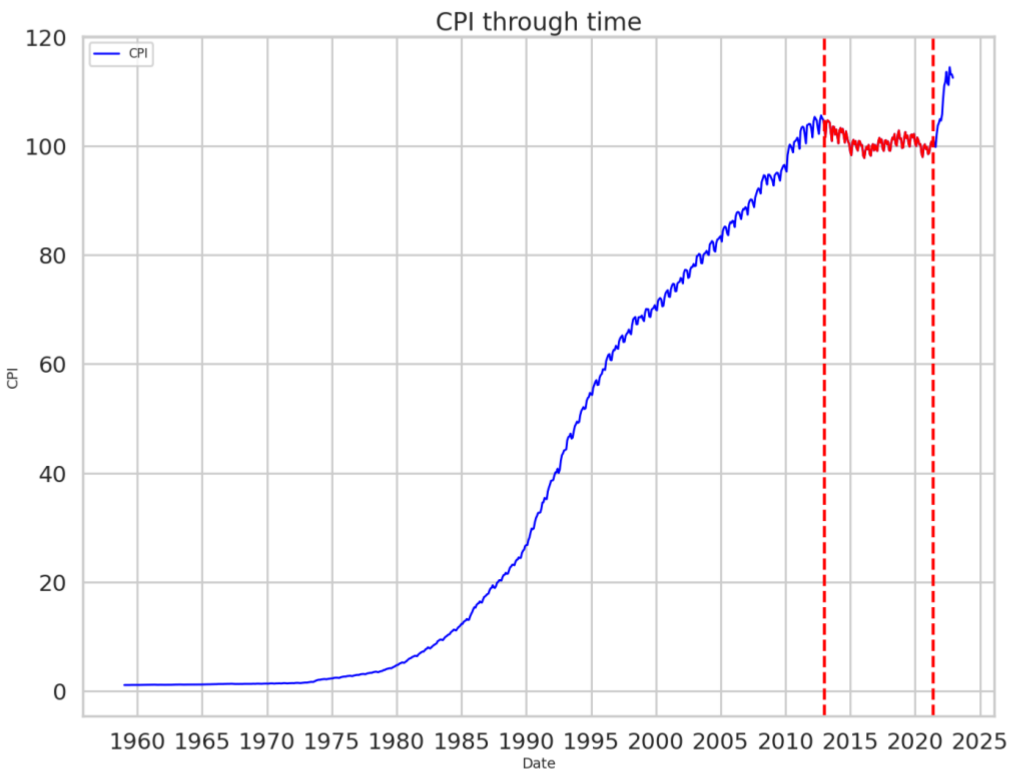
In general, this part of the report can be long, and time consuming, as univariate, and multivariate visualizations of all variables need to be generated along with descriptive statistics, however, as previously mentioned, it is an imperative step towards a good and parsimonious model.

Past Inflation Values

The first thing that needs to be done in exploring the data, is to visualize the CPI, the HICP and the basic sub-classes of the HICP (i.e., unprocessed food, services, energy etc.), as they constitute the most fundamental inflation measures. In doing so:

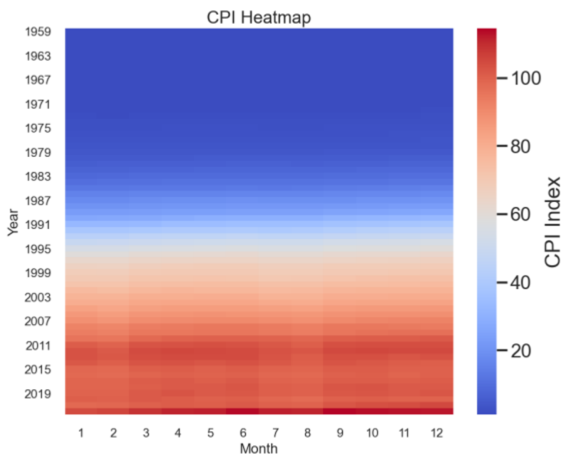
(a) the extent of the similarity of those two inflation indices will be clearer.
(b) the 'history' of Greek inflation will be illustrated.
(c) the relationship of each sub-class to inflation will be brought forward.

In general, CPI is the most widely used measure of inflation in Greece and the dataset consists of 768 CPI values that range from January 1959 to December 2022. CPI is calculated according to a base year, in which case is 2020 and the mean CPI for this specific year is 100. Therefore, it is eminent that due to the fact that CPI is a generally increasing measure in the long-run (see heat-map 1), the vast majority of values before 2020 score bellow 100 units (see appendix 6).

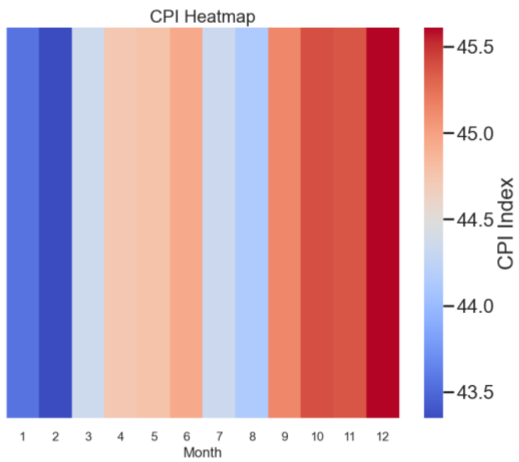


Line chart 2 - CPI trend from 1959 to 2022. Designed on python based on ELSTAT data.

Line-chart 2 depicts the values of the CPI from January 1959 to December 2022, which show a noticeable overall upward trend. The CPI values appear to increase rapidly and only stabilize after peaking in around 2012. The period between 2012 to 2022 seems to be a time of inflation stabilization in Greece, as indicated by the colored section of the line in line chart 2. One significant factor that may have contributed to the stabilization of prices in Greece, is the economic crisis that began in 2009 and lasted for over a decade, as highlighted by Elliott in 2019 (Elliott, L., 2019). However, from 2022 onwards, there is a spike in CPI values, likely due to the Russian invasion of Ukraine and the emerging energy crisis as reported by the Financial Times in 2022 (Financial Times, 2022). Additionally, some degree of seasonality appears to be present in the CPI values, becoming more apparent over time. This may be due to seasonal fluctuations in food items or energy sources, as well as other factors such as weather, holidays, or agricultural cycles (Mills, Patterson, 2011).
To investigate which months have the highest CPI values, the mean CPI for each month was calculated. It was found that January and February have the lowest mean CPI values (around 43.4), while October, November, and December have the highest CPI values, with December's mean CPI value being around 45.61 (as depicted in heat-map 2). Overall, the average CPI value is 44.64, and there are no null or outlier CPI values present in the cleaned and prepared dataset provided by the National Statistical Authority of Greece, thus ensuring its quality.

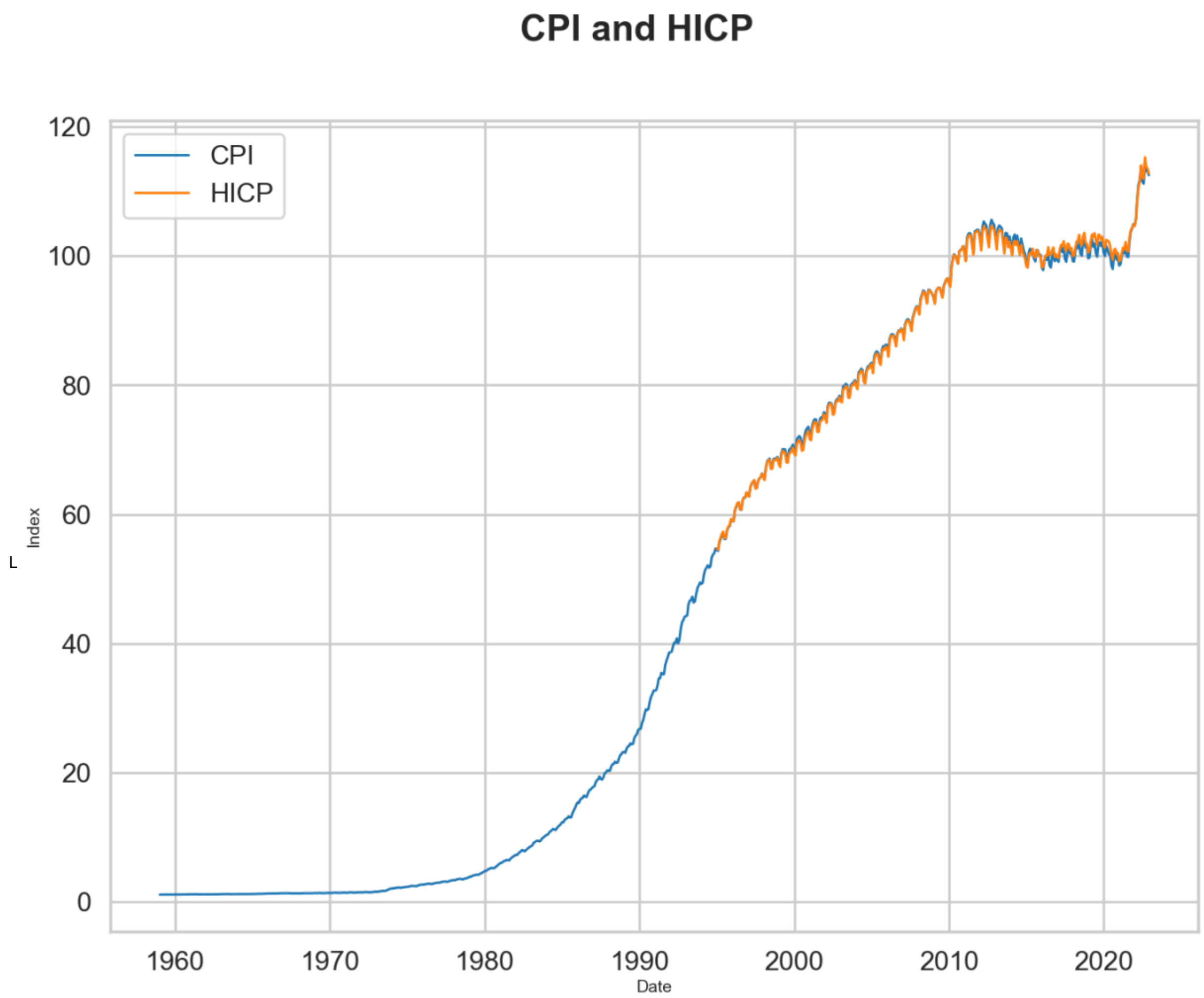


Heat-Map 1 - Annual and Monthly CPI values. Designed on python based on ELSTAT data.



Heat-Map 2 - Mean Monthly CPI values.Designed on python based on ELSTAT data.

Line chart 3, illustrates the trend lines of both CPI and HICP in one chart. It is clear that the HICP is very similar to CPI both in terms of magnitude and direction of changes, although there are minor differences visible in their respective trend lines that occur due to some of the underlying differences in HICP measurement and calculation as elaborated in section 2.2 – Measuring inflation. Additionally, it is important to highlight that the HICP only has available records from January 1995 to December 2022 (324 records). Hence, the number of records is less than half in comparison to the CPI and therefore, the selected target variable for this research project is the CPI. It should be mentioned that the CPI is not appropriate for inflation comparison with other countries, however, since the purpose of this research project is to compare different methods for inflation forecasting, it is deemed that using the CPI as target variable instead of the HICP is not out of scope.

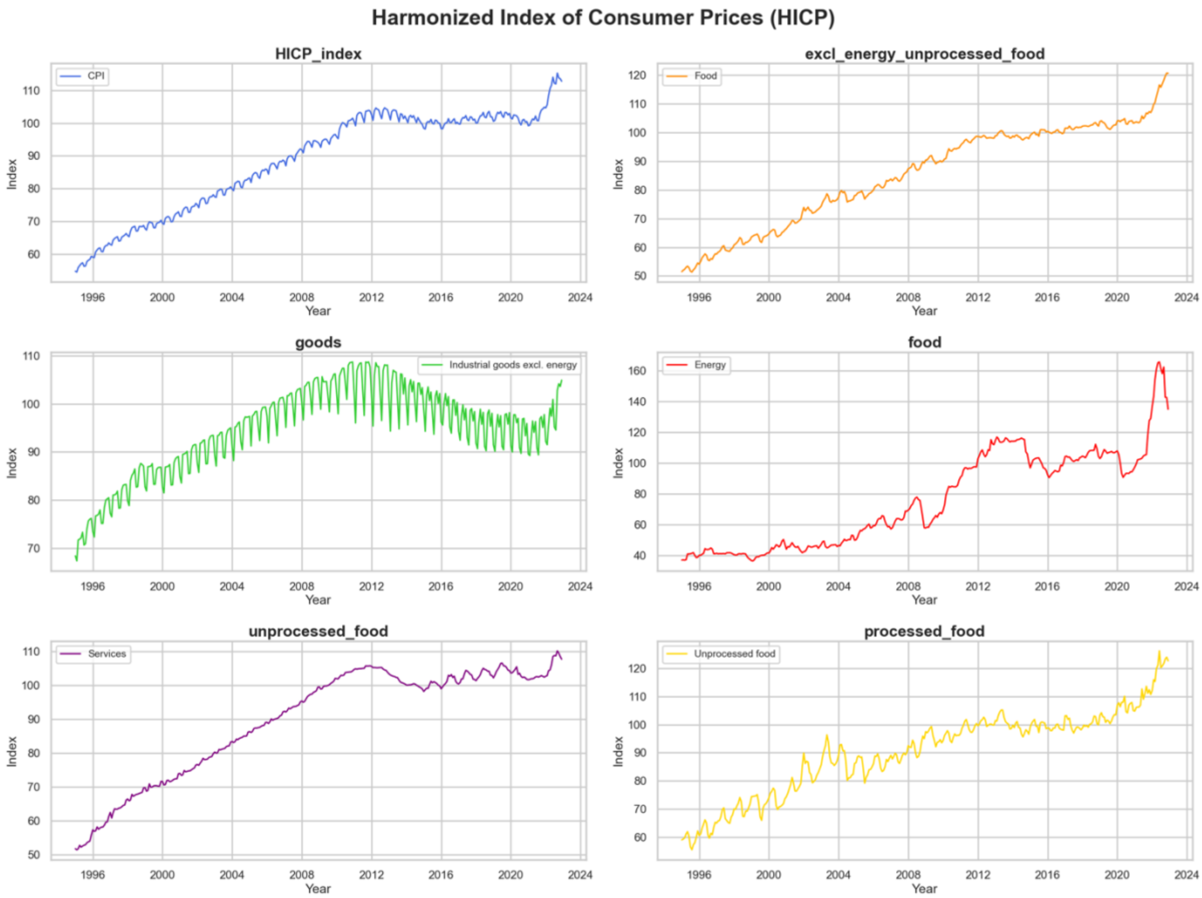


Line chart 3 - CPI and HICP trend lines. Designed on python based on ELSTAT data.

Dashboard 1 displays line charts of some of the basic sub-classes of HICP, which show that these sub-classes behave similarly to HICP in terms of index value spikes and declines, and at similar timings. However, it appears that some sub-classes, such as energy, food, and unprocessed food, contribute to pushing HICP higher, while others, such as services and industrial goods excluding energy, tend to pull HICP lower.

Of particular interest is the fact that industrial goods excluding energy, which historically has had the highest mean of all CPI sub-indices, tends to have lower values during periods of rising energy prices since energy prices are not included in this subclass. It is worth noting that energy and unprocessed food values seem to have risen dramatically around 2010 and after 2021, which may have been influenced by various factors such as the pandemic and the Russian invasion of Ukraine.

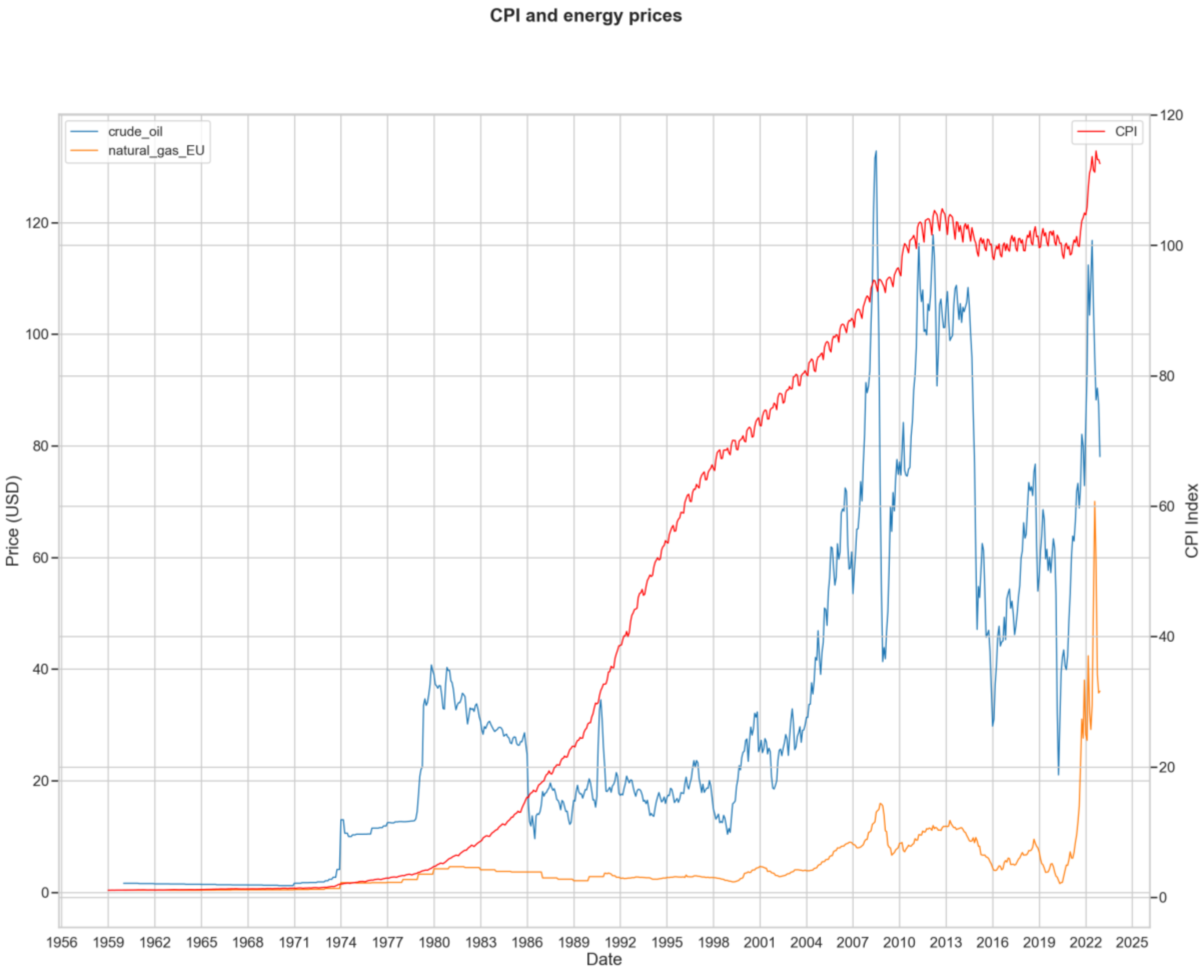
Moreover, industrial goods excl energy seems to exhibit high seasonality, which may also explain the high seasonality observed in both HICP and CPI values. Given these findings, it is clear that basic commodity prices, such as fossil fuels, foods, and crops, warrant further examination since food and energy subclasses of HICP appear to be important inflation drivers that are closely linked to basic commodity prices.



Dashboard 1- HICP sub-classes Trends from 1995 to 2022 and HICP Trends from 1995 to 2022. Designed on python based on ELSTAT data.

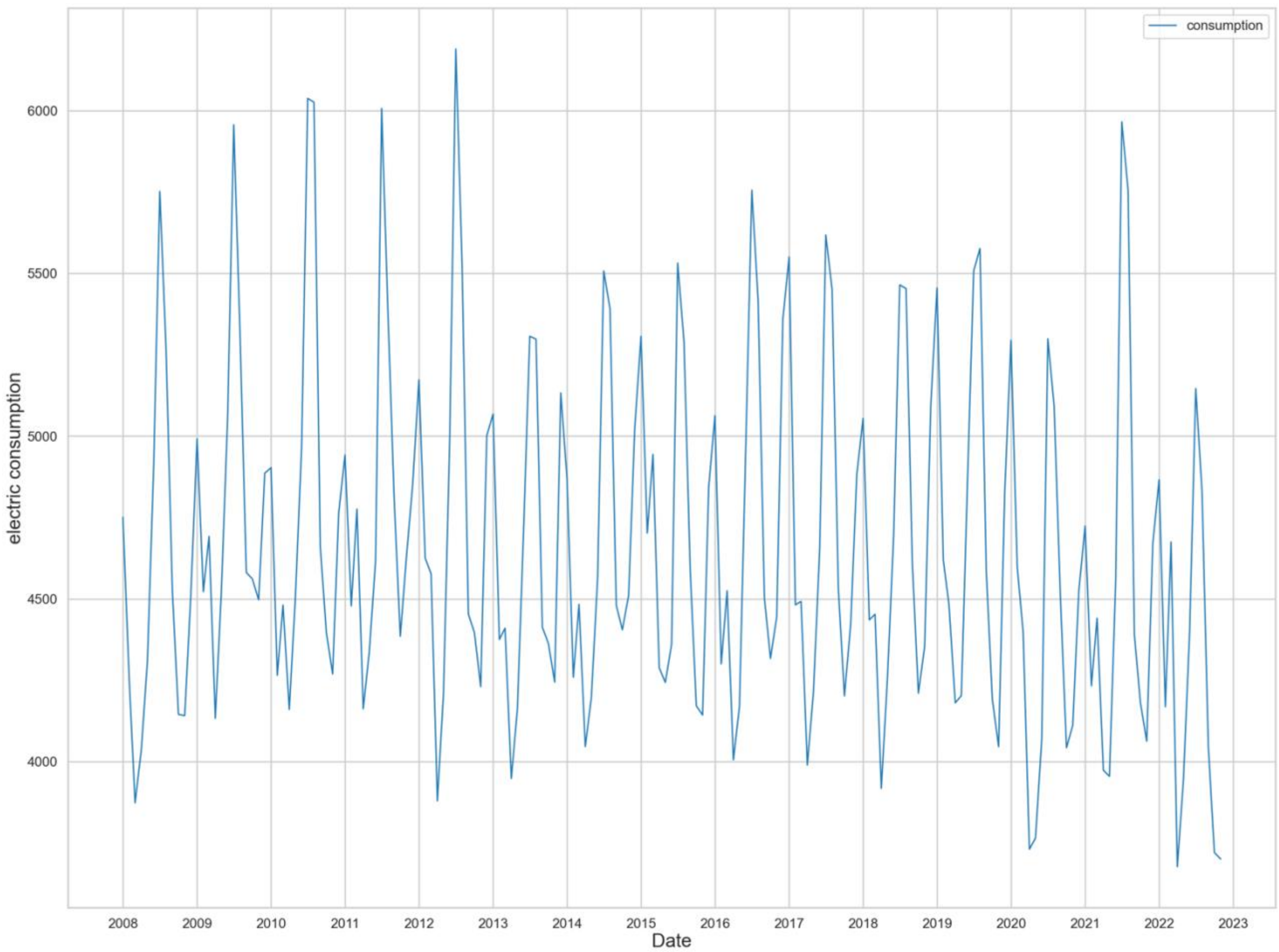
Inflation*, Energy,* Commodity Prices *& import prices*

To better understand the relationship between basic commodity prices and CPI, it is essential to visualize their historical trends. Line-chart 4, below, provides a graphical representation of the prices of crude oil and natural gas over time, with the use of a secondary axis to display CPI values simultaneously, allowing for an immediate comparison. The chart clearly shows that changes in crude oil and natural gas prices do have a significant impact on CPI values, as evidenced by the visible spikes in CPI values between 2007-2015 when energy prices surged, and again after 2020, when the rise in energy prices led to an increase in CPI values. However, it is worth noting that the relationship between basic commodity prices and CPI values is not as pronounced for earlier periods. Therefore, it is important to continue monitoring the relationship between basic commodity prices and CPI values, especially during times of energy price fluctuations and other economic disruptions that may affect commodity prices. By doing so, policymakers can make informed decisions about inflation management and economic stability.



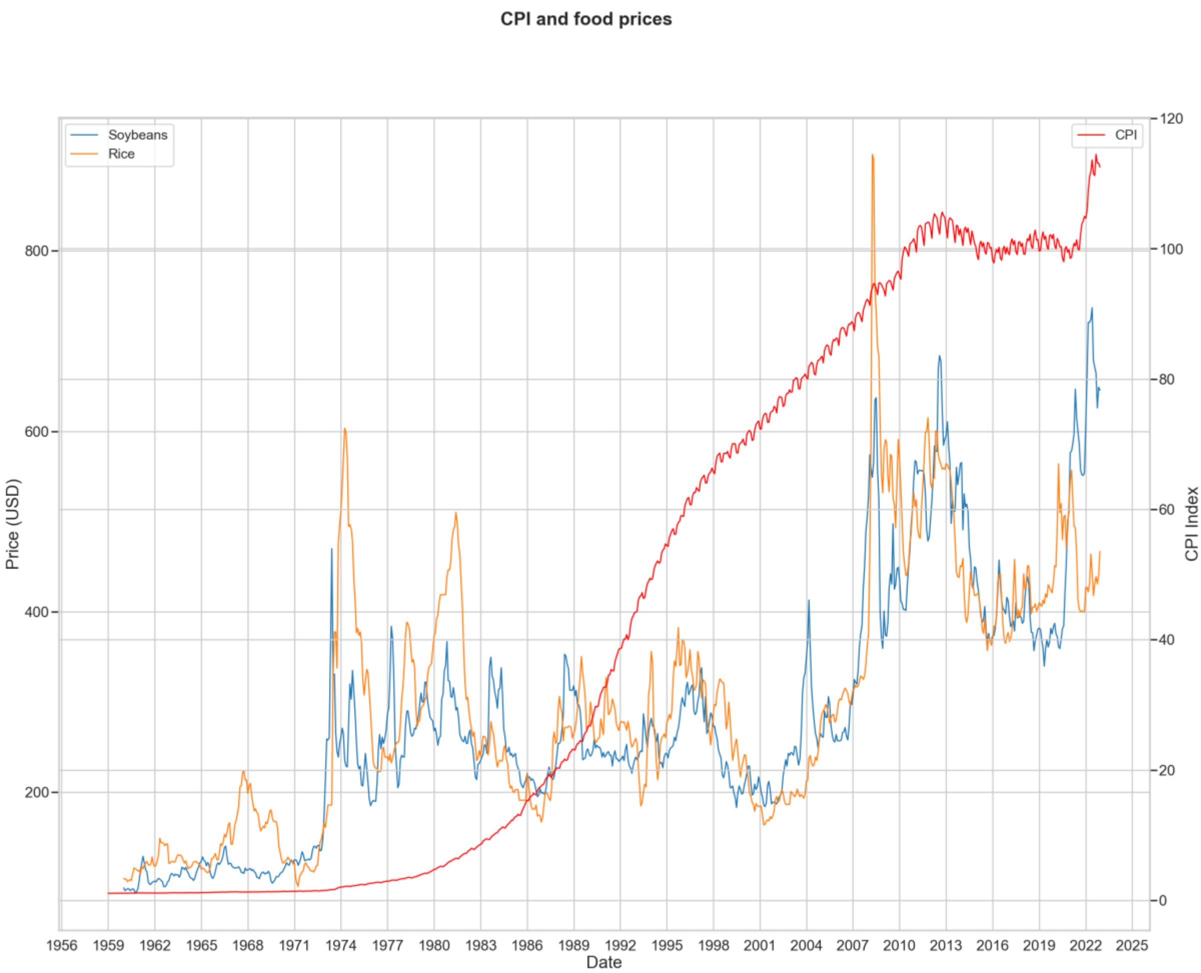
Line Chart 4- CPI and Energy prices trend lines. Designed on python, based on ELSTAT data.

Despite the usual relationship between oil and natural gas prices and energy consumption, it appears that energy consumption in Greece is not strongly correlated with either of these factors or with the CPI. After the 2010 Greek economic crisis and the 2020 COVID-19 pandemic, energy consumption in Greece has declined, but overall, it does not seem to have a direct correlation with CPI (see line chart 5). It is worth noting that electricity consumption in Greece exhibits high seasonality, with increased usage during colder months when electricity is used as a primary source of heating. This suggests that other factors, such as domestic energy production, energy policies, and consumption patterns, may also play a role in shaping energy prices and their impact on CPI. Therefore, a comprehensive analysis of various factors affecting energy prices and consumption in Greece would be necessary to better understand their relationship with CPI and overall inflation dynamics in the country.



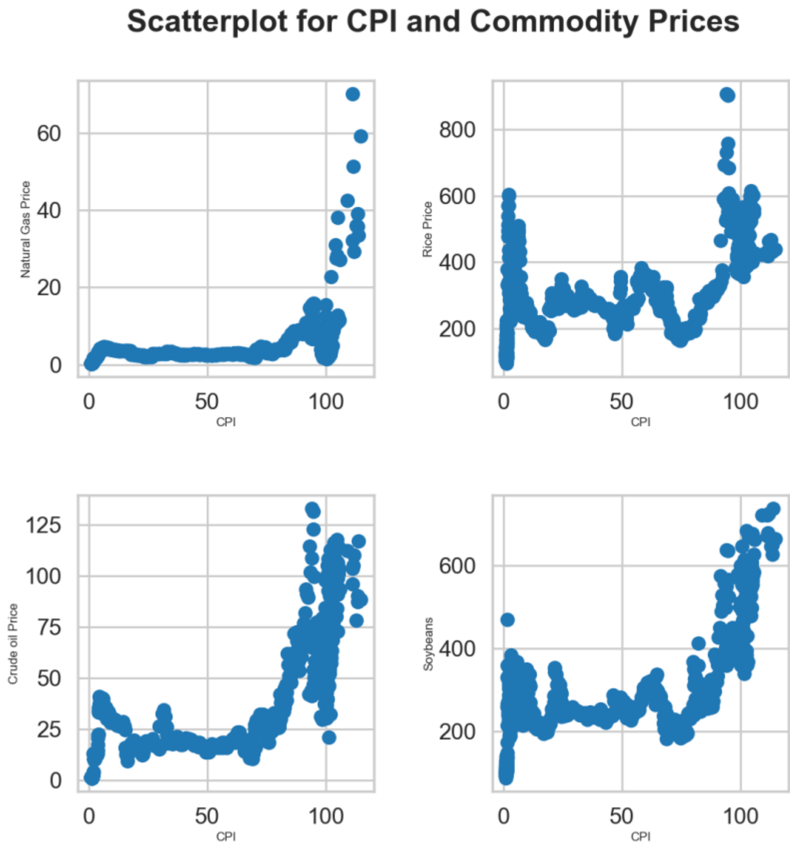
Line Chart 5- Electric consumption in Greece in MWh from 2008 to November 2022. Designed on Python, based on ELSTAT data.

Moving on, line-chart 6, bellow, presents the trend lines of Soybeans price and Rice price and a secondary axis presenting the CPI values at the same time period. The lines are rather similar to those of line chart 4 in terms of trajectory, and there is no eminent relationship for values before 2007. However, during periods of soybeans and rice price spikes, CPI also presents significant increases. After the 2007-2013 increase in wheat and rice prices, prices seem to drop. CPI values follow suite, and also seem to drop after 2013.



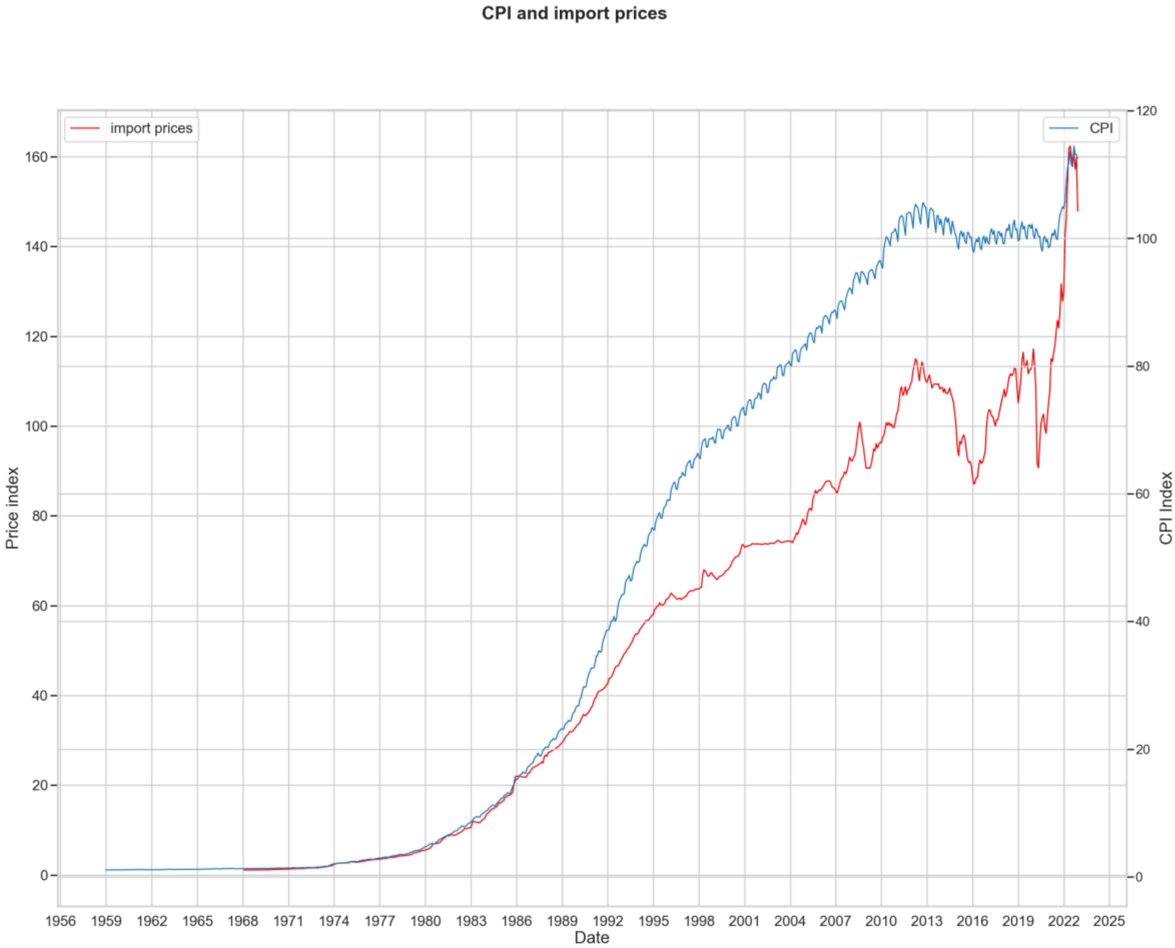
Line Chart 6- CPI and basic commodity prices trend lines. Designed on python based on World Bank data.

The relationship between basic prices and CPI can be further demonstrated using scatterplots that show crude oil, rice price, soybeans price and natural gas price on their x axis and CPI prices on their y axis. In this manner, if a pattern is recognized, that means that there is some sort of relationship between the variables. Scatterplot 1 demonstrates that there is no eminent relationship between CPI and commodity prices for lower CPI values, but when the CPI is high enough (above 80) it is eminent that commodity values are also high.



Scatterplot 1- CPI and commodity prices. Designed on Python based on ELSTAT data.

Line chart 7, bellow, presents the trend lines of CPI and import prices index in a dual axis. It is eminent that there is a strong relationship between import prices and CPI. The CPI line and the Import prices line move towards the same direction and at a very similar timing, although import price rises and declines occur a few weeks earlier than those of CPI. The linear correlation between CPI and import price is also extremely high with a Pearson correlation coefficient of 0.98. It is hence suggested that import prices might be a good inflation predictor.

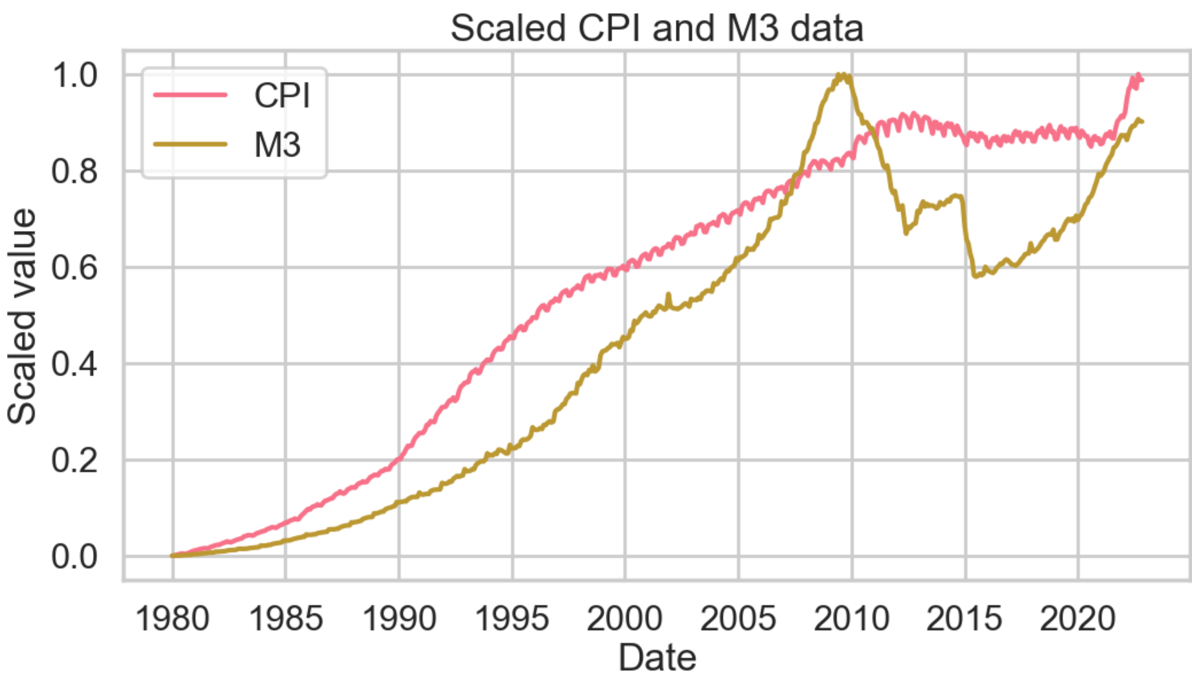


Line Chart 7- CPI and import prices trend lines. Designed on python based on ELSTAT data.

Inflation and Money Quantity

As explained is section 2 – literature review the quantity theory of money was one of the first theories that endeavored to explain the underlying causes of inflation. It suggested that as money stock increases, inflation rises, and as money stock decreases, inflation tends to decline. It is therefore imperative, that a bivariate visualization of a line chart that illustrates the trend lines of CPI and money quantity over time is designed, in order to explore any potential relationships between CPI and money stock as suggested by the quantity theory of money. In line chart 8, CPI and M3 (money stock) trend lines have been generated and some interesting findings should be noted. It shouldn't be omitted to highlight that for reasons of comparability, both CPI and M3 have been scaled to only take values between 0 and 1, using the MinMax scaler.

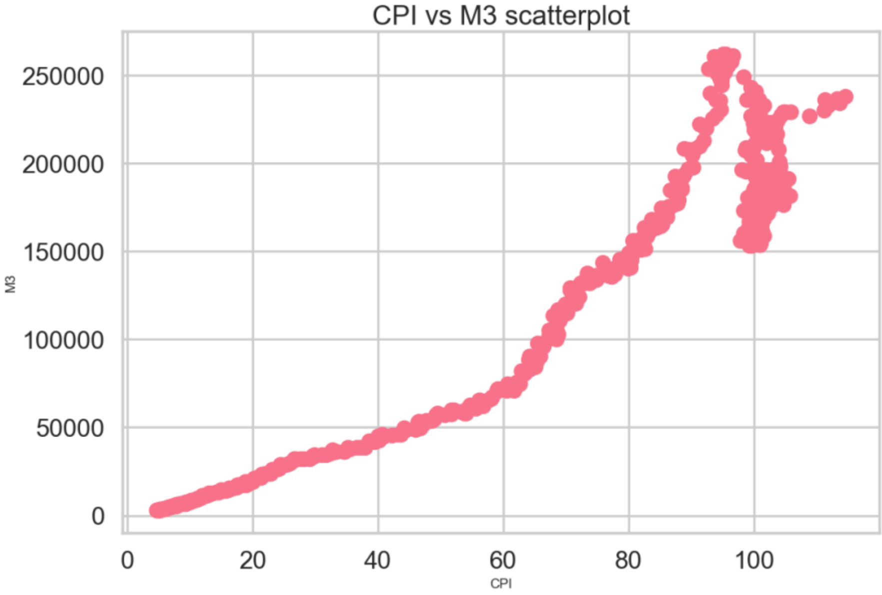
First, both lines present an overall upward trend. Additionally, CPI and M3 values appear to rise and decline at very similar timings. For instance, M3 had been rapidly increasing from 2008 until 2013 (during the Greek financial crisis). Soon after the rapid increase in M3, a significant spike in the CPI values is also observed (2010- 2015). Another example is that M3 values rose significantly after 2020. CPI followed suite and inflation also rose. It should be highlighted however, that CPI value changes delay to appear in comparison to M3 value changes in terms of months or even years. This might occur for many reasons, as explained in section 2-literature review such as: asset price inflation, the fact that the velocity of money and average prices are not included in the generation of the graph (as indicated by the equation of exchange), or simply because the market is reluctant to price increases (especially during periods of crisis, such as the pandemic or the 2010 Greek financial crisis). However, this lagged relationship between M3 and CPI could be taken advantage of, and be used for inflation forecasting.



Line Chart 8- scaled CPI and M3 trend lines. Designed on python based on ELSTAT data.

In general, there appears to be a strong relationship between CPI and M3, and this is also justified by both the theoretical background as elaborated in the Quantity theory of money section of the literature review and by scatterplot 2. It is clear in scatterplot 2, that as M3 rises, CPI also tends to rise and when M3 drops, CPI tends to decline. Although the relationship is not linear, the Pearson's correlation coefficient seems to come into agreement with the above observations with a correlation score of 0.95.

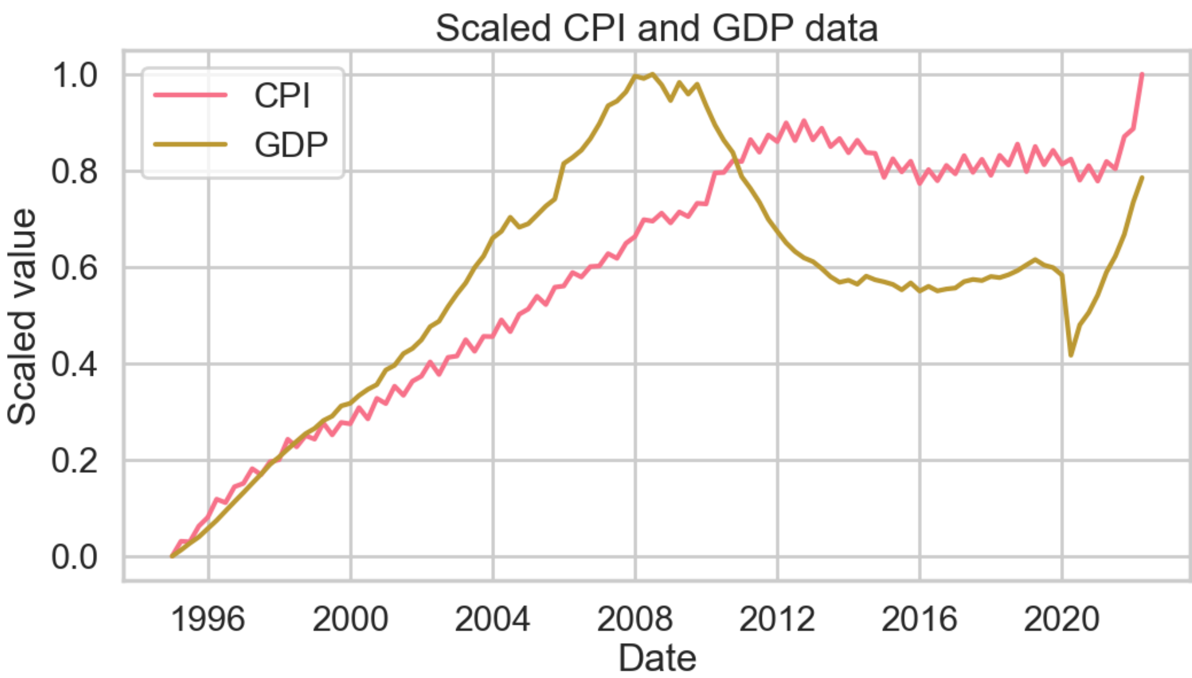
Overall, M3 seems like a valuable predictor of inflation. In terms of data quality, the money quantity dataset (M3), has been collected by ELSTAT, the Greek government's official statistical authority, that has cleaned and prepared the dataset, and hence its quality is out of question. However, it should be noted that M3 values only date back to January 1980 and hence, the dataset only consists of 504 values.



Scatterplot 2- CPI and M3 scatterplot. Designed on python, based on ELSTAT data.

According to the quantity theory of money, the product of the money stock and the velocity of money, equals the product of average prices and total value of goods (GDP) as formulated by Fischer (Fisher, I., 1912). Therefore, although the velocity of money is not a measurable quantified macroeconomic concept, GDP is one of the most basic macroeconomic measurements and its relation to CPI should be illustrated. It should be noted, that for reasons of comparability, both measurements have been scaled to take values between 0 and 1, using the MinMax scaler. In line chart 9, GDP seems like a ‘backward-shifted’ CPI curve, in the essence that, spikes and declines in the CPI curve follow a very similar pattern to that of the GDP curve with the only difference that the timing of the changes in the values of CPI is delayed.

More specifically, until around 2010 both lines present an upward trend. However, GDP presents a radical spike in its values around 2008 and a dramatical decline in its values after 2010. CPI values follow suite and present a significant spike that peaks around 2012 (4 years later) and declines after 2013. Similarly, at the end of 2020, a significant spike in the GDP values is visible, while CPI values also seem to eminently increase after 2022. This relationship between CPI ang GPD may suggest that GDP constitutes an important CPI predictor, something that is further reinforced by the eminent delay of CPI values’ changes in contrast to GDP. It seems like that the exploratory analysis of the dataset at hand, comes in agreement with the theoretical suggestions elaborated in section 2- literature review. Additionally, in terms of data quantity, GDP has been also measured by the Hellenic statistical authority and hence it is considered a trustworthy and integral dataset (as also indicated by the fact that there are no missing values or incorrectly recorded values). Finally, in terms of data quantity, the dataset is somewhat problematic. First of all, the data only dates back to 1995, while at the same time, GDP has only been quarterly measured. This decreases the number of records to only around 112 data points.

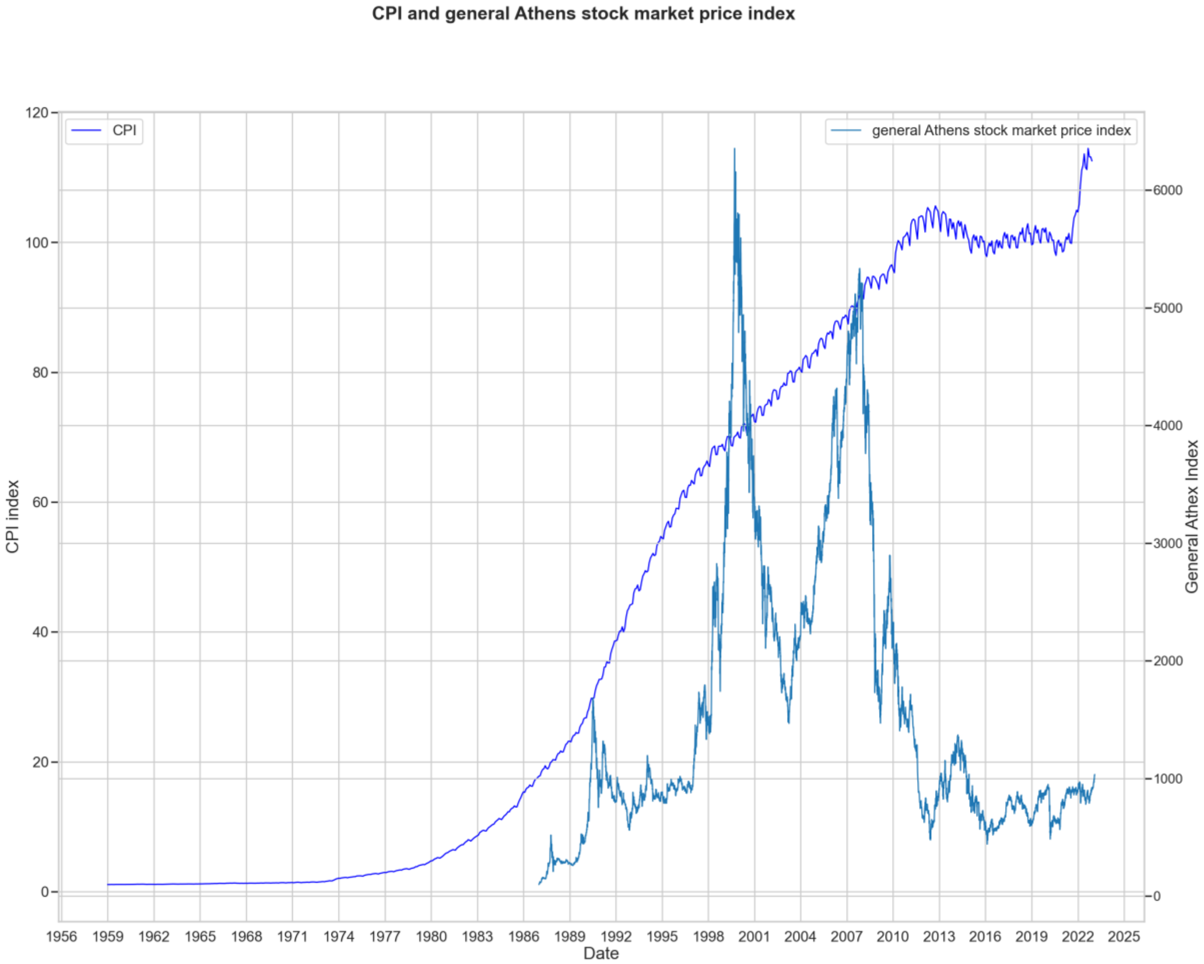


Line Chart 9- scaled CPI and GDP trend lines. Designed on python, based on ELSTAT data.

Inflation and Asset prices

Another big chapter in inflation forecasting is that of asset prices. As explained previously, money stock is an important predictor of inflation, when it is injected into the real economy. However, money expansion may also occur in asset markets such as in purchases of stocks, bonds, companies, or houses. When this happens, money is not injected into the real economy directly and hence CPI values may delay following the M3 increase, as shown in line chart 8 where the relationship between CPI and M3 is illustrated. In an endeavor to understand how CPI is related to asset prices, several charts have been designed using Python's Matplotlib library. Firstly, line chart 10, presents the trends of CPI values and avg. Athens stock market prices index through time in a dual axis, so that the indices are easily comparable.

It is eminent, contrary to what the theory suggests, that there is no clear relationship between CPI and avg. stock market prices, as avg. Athens stock market prices follow an overall downward trend while CPI (and money supply) follows an upward trend. It is important to note however, that between 2013 and 2021, when CPI is stabilized, avg. stock market prices are also to some extent stabilized. However, avg. stock market prices do not follow the upward trend that is eminent in CPI line after 2022 and in general the magnitude of fluctuations is different. Additionally, the quantity of stock market data is extensive in comparison to other macroeconomic datasets, as it dates back to January 1987 and is recorded on a daily basis. The quality of the data is not questionable as it has been collected and prepared by the Bank of Greece.

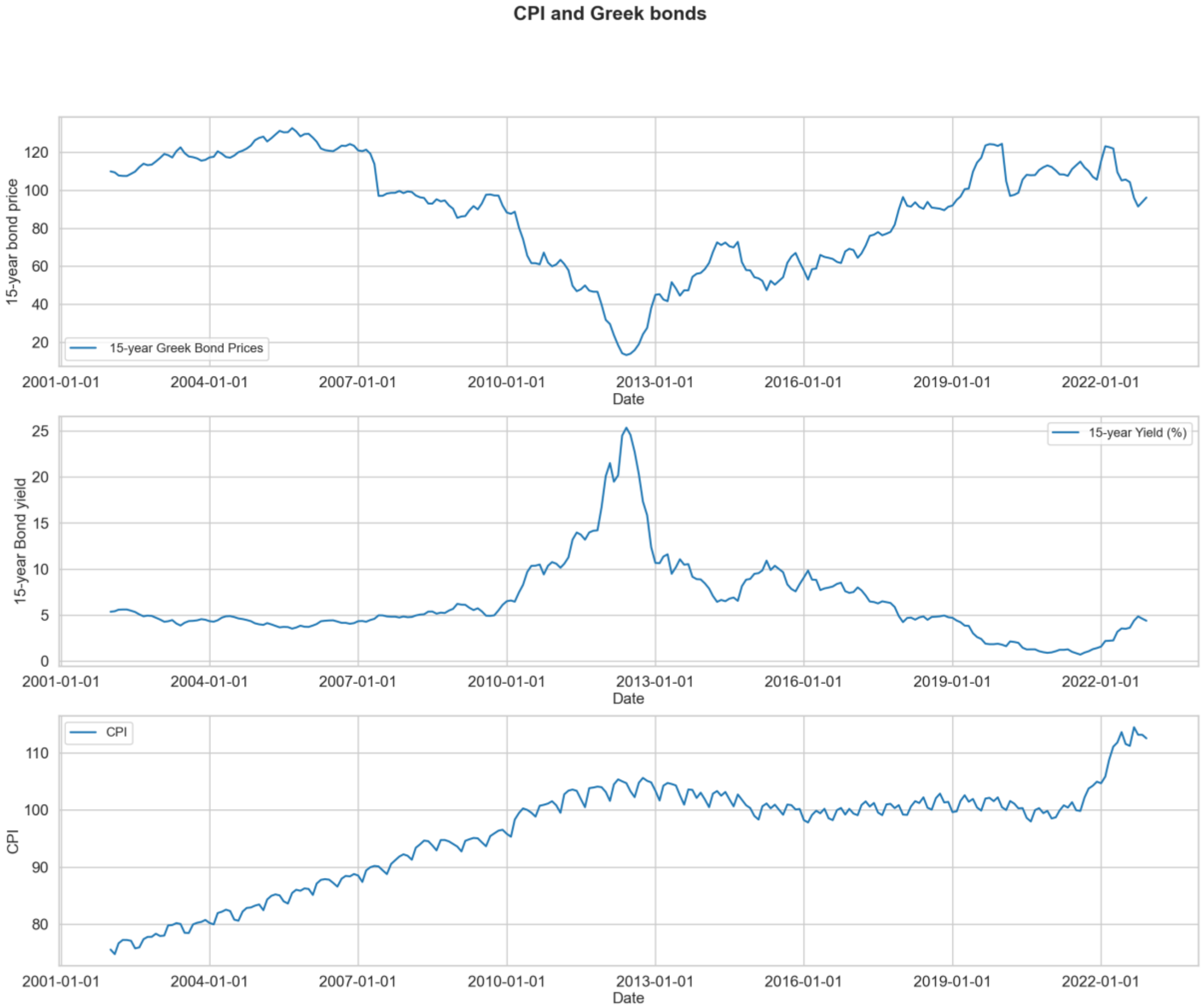


Line Chart 10- CPI and general Athens stock market index trend lines in dual axis. Designed on python, based on ELSTAT & Central Bank of Greece data.

The relationship between bond prices, bond yields, and CPI has been extensively studied in the literature. According to Choudhry (Choudhry, 2013), bond prices and yields are inversely related, with higher bond prices leading to lower yields and vice versa. This relationship arises because bond prices represent the present value of the bond's future cash flows, while bond yields represent the effective interest rate that the investor earns on the bond. Thus, when the bond price increases, the yield decreases, and when the bond price decreases, the yield increases.

In contrast, the relationship between bond yields and CPI is positive. As Ang and Piazzesi note (Ang, Piazzesi, 2003), inflation affects the purchasing power of a bond's fixed interest payments. Therefore, when inflation rises, bond yields also rise to compensate investors for the decrease in the value of the fixed payments. Conversely, when inflation falls, bond yields decrease to reflect the decreased risk of inflation eroding the value of the fixed payments.
In summary, the relationship between bond prices, bond yields, and CPI is complex and interrelated. Bond prices and yields have an inverse relationship, while bond yields and CPI have a positive relationship. These relationships reflect the impact of inflation on the value of fixed interest payments and the present value of future cash flows.

These relationships are eminent in line chart 11, bellow, where -although with bigger fluctuations – 15 year-Bond yield seems to be related to CPI, while bond price seems to be negatively related to both. However, both the price and the yield of the 15-year Greek bond, might be valuable predictors in forecasting future inflation values. Overall, the quality of the data of both the bond price and the bond yield, is high, and the dataset is cleaned and organized as it was supplied by the Central Bank of Greece. However, the data only date back to 2001 and thus the quantity of the data may not be sufficient for this research project’s purposes.



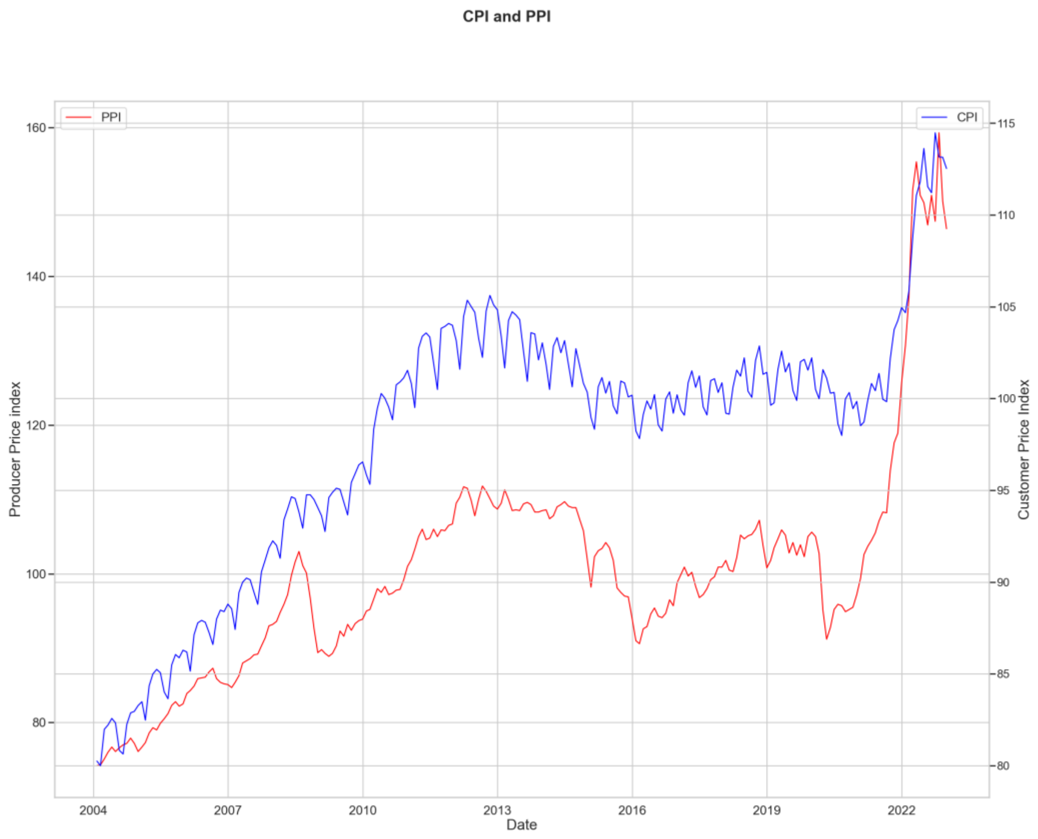
Line Chart 11- CPI, 15-year bond price and 15-year bond yield. Designed on python, based on Central Bank of Greece data.

Inflation and Industry

The relationship between CPI and PPI is an important chapter in macroeconomics as both measures are used in conjunction to measure inflation in economies. While the CPI measures the price changes in a basket of goods and services, PPI measures the price changes in the scope of goods and services that businesses produce. Many studies have shown that there is a significant relationship between CPI and PPI, but the direction and the magnitude of the relationship depends on many factors affecting the economy. For example, a study conducted in 2012 in the US discovered that the relationship between PPI and CPI is very strong in the long run but not always immediate (Lashgari et al., 2012).

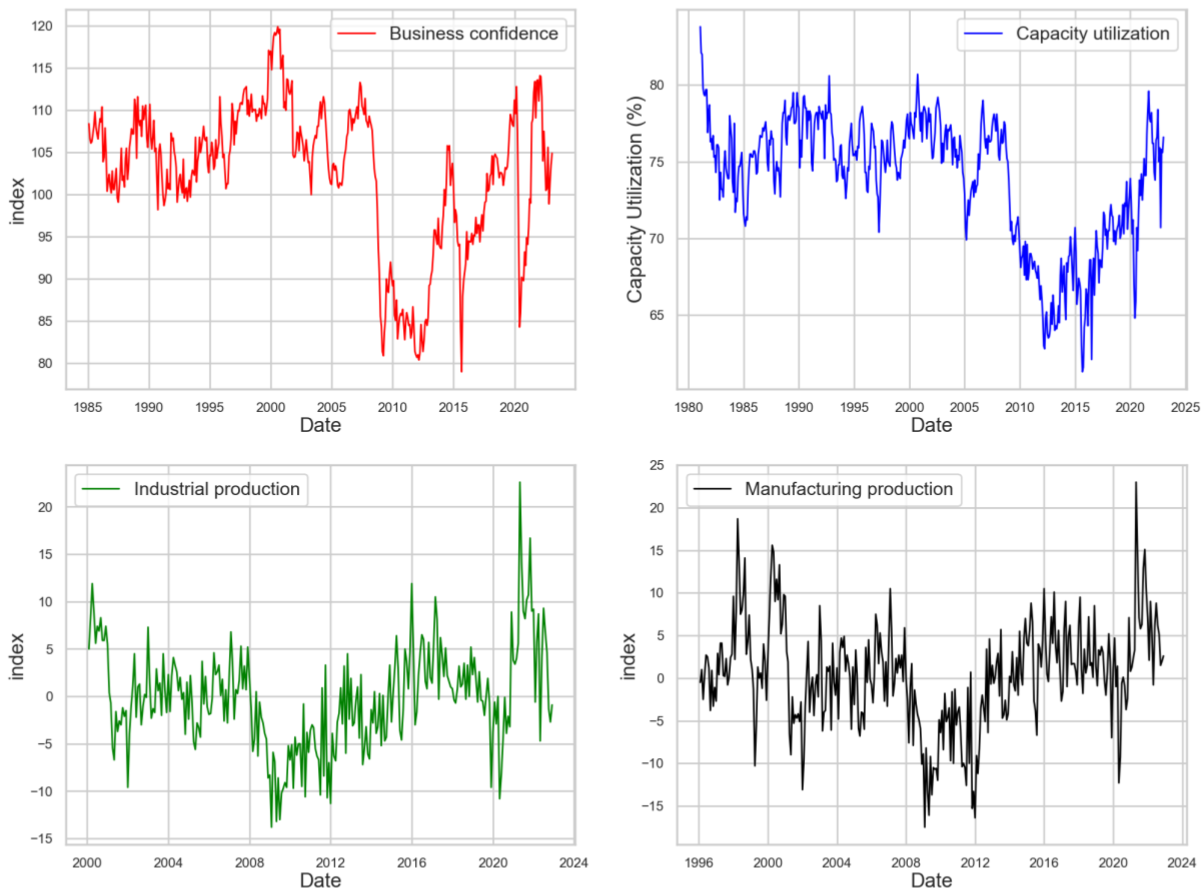
Other studies suggest that CPI and PPI are affected by many factors such as exchange rates, global economic conditions and changes in supply and demand. A notable research is that of Rautava and Vilmi conducted in 2017, suggesting that changes in the exchange rates impact to a significant extent the relationship between CPI and PPI, as prices of imported goods are affected (Rautava and Vilmi, 2017).

In line chart 12 it looks like PPI and CPI are indeed related and that is supported by the similar trajectory that both PPI and CPI lines follow and at a very similar timing. However, it should be noted that after 2020, the magnitude of increase in PPI values is bigger than that of the CPI values and this may occur due to the increased import prices that the pandemic brought forward. In general, the dataset quality is high as it has been cleaned and prepared by the Hellenic statistical authority (ELSTAT). However, it should be noted that PPI values are recorded in Greece only from 2004 onwards, and hence, the quantity of the data might not be sufficient for this project’s purposes.



Line Chart 12- CPI and PPI. Designed on python, based on ELSTAT data.

It has been suggested that factors such as business confidence, capacity utilization, industrial production and manufactory production may be linked to the CPI. Higher levels of business confidence may indicate that businesses expect increased sales and hence this may lead to increased economic activity and potentially higher prices. For instance, a 2014 study showed that business confidence is an important predictor of inflation in the Euro area (Belke, Klose, 2014). Capacity utilization may also be a significant inflation predictor as high-capacity utilization may be an indicator of high demand which can contribute to higher prices (Stracca, L., Terlizzese, D., & Bonci, R., 2013). Similarly, industrial production and manufacturing production may be closely related to CPI as increased production may be translated to increased demand and hence, higher prices. It is eminent from line chart 13 that all indices behave similarly and follow the trajectory of the CPI trend line. However, using variables that correlate with each other in ML models may lead to issues with multicollinearity. The linear correlation of the industrial indices can be tested by calculating the Pearson correlation coefficients of all the indices. Table 3 shows that capacity utilization and business confidence are strongly correlated, while manufacturing production and industrial production are also very strongly correlated. The quality of all data is high as they have been collected and prepared by the Hellenic statistical authority (ELSTAT). However, capacity utilization has a wider range of data in comparison to business confidence, and manufacturing production also has a wider range of data in comparison to industrial production. At the same time due to high correlation coefficients between the two couples of variables, it is suggested that only capacity utilization and manufacturing production will be potential candidates for future models to avoid issues with multicollinearity.



Line Chart 13- Business Confidence, Capacity utilization, Industrial production and Manufacturing production. Designed on python, based on ELSTAT data.

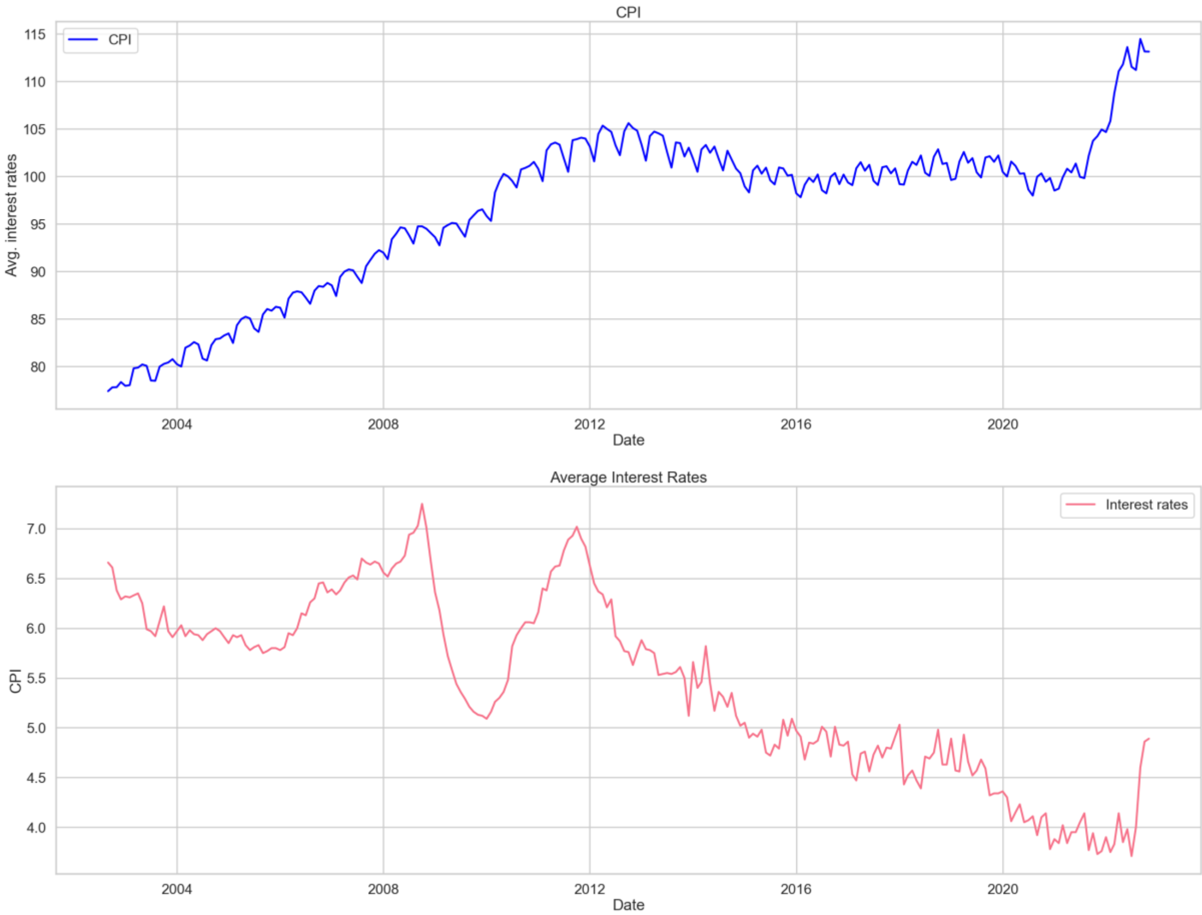
	business_confidence	capacity_utilization	manufacturing_production	industrial_production
business_confidence	1.000000	0.784779	0.514728	0.466638

	business_confidence	capacity_utilization	manufacturing_production	industrial_production
capacity_utilization	0.784779	1.000000	0.237004	0.282037
manufacturing_production	0.514728	0.237004	1.000000	0.914440
industrial_production	0.466638	0.282037	0.914440	1.000000

Table 3 – Pearson's correlation coefficients of industrial indices

Inflation and National economy

Moving on, as explained in section 2 of this report, interest rates are closely related to the CPI as countries adjust interest rates to influence inflation. It is hence expected that the lines of the CPI and average interest rates move towards the same direction and at a similar timing. However, line chart 14 shows that the average interest rates trend line only follows the trajectory of the CPI trend line at high inflation periods. This is eminent during the 2008-2013 period when CPI values, similar to average interest rates values, peak before declining, and even more eminently in 2022 when the spike in CPI values is followed by an increase in the average interest rates. In general, interest rates are used to adjust inflation rates and are used as a remedy after inflation spikes are observed, and hence, it is not clear whether interest rate measurements are capable of predicting future inflation values, especially in the short term. Additionally, the quantity of data available in this dataset is quite small, but the quality of the data is high, since it has been cleaned and prepared by the National Bank of Greece.

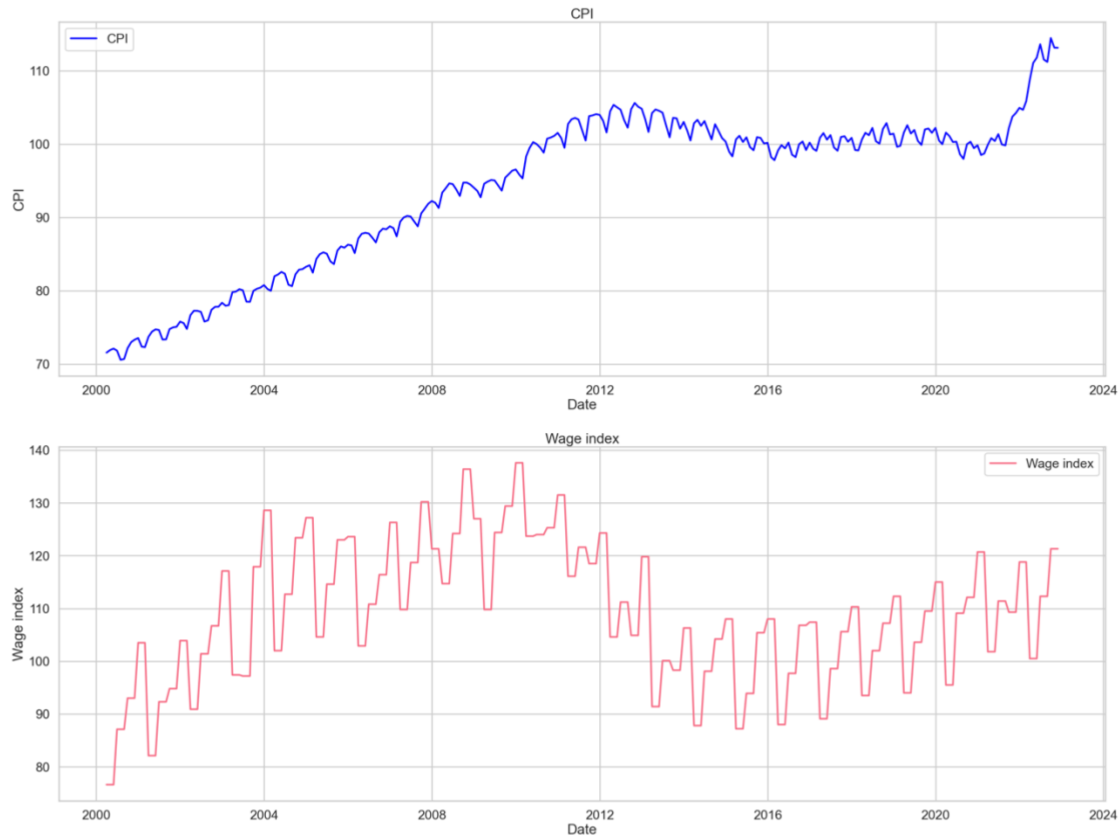


Line Chart 14- CPI and Average interest rates trend lines. Designed on python, based on Central Bank of Greece data.

Another important aspect of national economy is the wage index which tracks the changes in wages during a period of time and illustrates the growth and trends of wages (Mishkin & Serletis, 2011).

In line chart 15, it seems like CPI values lag behind wage index values, however, the increases and declines and the overall trend of the lines are towards the same direction. It is important to mention however, that the wage index seems to be highly seasonal and that may occur due to the large amount of people being occupied in the tourist industry during the summer season in Greece. It should not be omitted that the quality of the dataset is also very high as it has been prepared by the Hellenic statistical authority. However, only quarterly data are available, and they have been transformed into monthly values via the forward-fill method. This explains the reason why the wage index trend line seems to move stepwise over time. It is significant to highlight then, that due to this transformation, the data may not be ideal for inflation forecasting and their original quantity is very limited, as there are only quarterly records from mid-2000 to 2023 (only around 90 records).

In summary, while the line chart illustrates a general positive trend and alignment between the wage index and CPI, the presence of seasonality, the transformation of quarterly to monthly data, and the limited size of available data should be acknowledged to ensure a comprehensive understanding of the observed relationship.



Line Chart 15- CPI and transformed wage index trend lines. Designed on python, based on ELSTAT data.

3.3 Data preparation

The third crucial step according to the Cross-industry Standard Process for Data Mining (CRISP-DM) methodology is that of data cleaning. This part of the methodology involves addressing issues such as missing values, outliers or other data quality problems. More specifically, data cleaning includes:

Data profiling, meaning a comprehensive understanding of the dataset's characteristics, such as variable types, distributions, and missing values identification (Batini, C., et al, 2009).

Handling missing values, meaning, deleting records with missing values or imputing missing values or even more advanced techniques like multiple imputation (Little, R. J., & Rubin, D. B., 2014).

Detecting outliers, meaning extreme observations that can affect the results of the analysis. Methods that can detect outliers include statistical approaches, clustering-based methods, and ML algorithms (Hodge, V., Austin, J., 2004).

Finally, the last part of the data cleaning process is ensuring the integrity of data, meaning ensuring that data is validated against predefined rules, cross-checking data across different sources and transform the data accordingly in order to resolve the inconsistencies.

In general, all the datasets have been retrieved either from the Hellenic Statistical Authority (ELSTAT), European Statistical Authority (EUROSTAT), the Bank of Greece or the European Central Bank (ECB) and hence, they have been thoroughly cleaned and prepared by the aforementioned authorities. However, even if this is the case, data profiling is detrimental before proceeding to the modelling phase of this report, as although outliers and missing values might not be present, consistency amongst the different variable data types and especially ensuring the timeliness of the data is imperative before generating any kind of models.

3.3.1 Data Profiling and Cleaning

Past inflation values

The CPI and the HICP data have been collected by the Hellenic statistical authority and the European statistical authority. It is important to compare the CPI to the HICP in order to find out which of the two measures of inflation is more suitable for being this report's target variable. After generating descriptive statistics, the HICP includes less values than the CPI that only date back to January 1995 (336 values) in comparison to CPI values that date back to January 1959 (768 values). It is also found that 25% of CPI values are bellow 2.294 while 50% of the values score bellow 32.7 and 75% of the values are bellow 88.49. Therefore, due to the exponentially increasing CPI values especially until 2010, it makes sense that the standard deviation of the dataset is quite large, and around 41.5. Because the HICP values are more recent in comparison to the CPI, the mean HICP value is 88.5 with a standard deviation of 15.4. 25% of the values are bellow 74.9 while 50% of the values are bellow 94.4 and 75% of the values bellow 101.3. This occurs because CPI values recorded before 1980 were extremely small in comparison to values after 1995, when the HICP was firstly recorded. It is eminent from line chart 2 that the HICP behaves similarly to the CPI in terms of growth rate, spikes, decline and timing. This is also confirmed by the Pearson correlation coefficient between the CPI and the HICP which scores 0.999. Therefore, since both the HICP and the CPI measure inflation very similarly and since CPI has a wider range of values (dating back to January 1959), the CPI will constitute this report's target variable. Again, HICP has been prepared by EUROSTAT and the quality of the dataset is excellent.

It is important to mention at this point, that the models that are generated aim at forecasting future CPI values in a time horizon of one month. The difference between CPI and inflation rate is that inflation is calculated by comparing the rate of change between the forecasted CPI value and the corresponding value in the previous year. For both the HICP and the CPI dataset, date is set as the index. Neither the CPI dataset nor the HICP dataset includes null values.

Energy, Commodity Prices & import prices

Data about basic commodity prices have been retreved from the World Bank and hence, the data have been collected, prepared and cleaned by financial experts. Hence, the quality of the dataset is out of question.

The dataset includes 69 different variables that illustrate the values of different commodity prices. However, only specific commodities are considered relevant to this research project. More specifically, for purposes of timeliness, variables that do not date back to January 1960 are taken out of the dataset. Moreover, Sunflower oil, Rapeseed oil and Banana (Europe), have many unregistered values (over 400), and hence they are also ruled out of the dataset.

Additionally, to avoid issues of multicolliniearity pearson's correlation coefficient has been calculated for all variables included in the dataset (see appendix 3). For each couple of variables that are strongly related (coefficient over 90%), the variable that is less relevant to the European and Greek economy is also rulled out. Hence, all crude oil prices, Iron ore, Copper and Platinum, except of the average crude oil price are deleted. Additionally, all Tea prices except of the avg Tea price are deleted. Groundnut oil, soybean oil, soybean meal, Maize, Wheat (US HRW), and Palm oil, are also deleted because they are highly correlated to Soybeans. Beef, chicken, Gold and Tobacco are taken out of the dataset due to their high correlation with Banana (US). The remaining variables are presented in table 4.

Commodity Variables		
Crude oil, average	Natural gas, Us	Natural gas, Europe
Cocoa	Coffee Arabica	Coffee. Robusta
Tea, avg 3 auctions	Coconut oil	Soybeans
Barley	Sorghum	Rice, Thai 5%
Shrimps, Mexican	Banana, US	Orange
Sugar, world	Sugar, EU	Sugar, US
Cotton, A Index	Logs, Malaysian	Sanwood, Malaysian
TSP	Rubber, RSS3	Phosphate rock
Aluminum	Urea	Potassium chloride
Nickel	Lead	Tin
Zinc	Silver	

Table 4 – Remaining Commodity prices variables

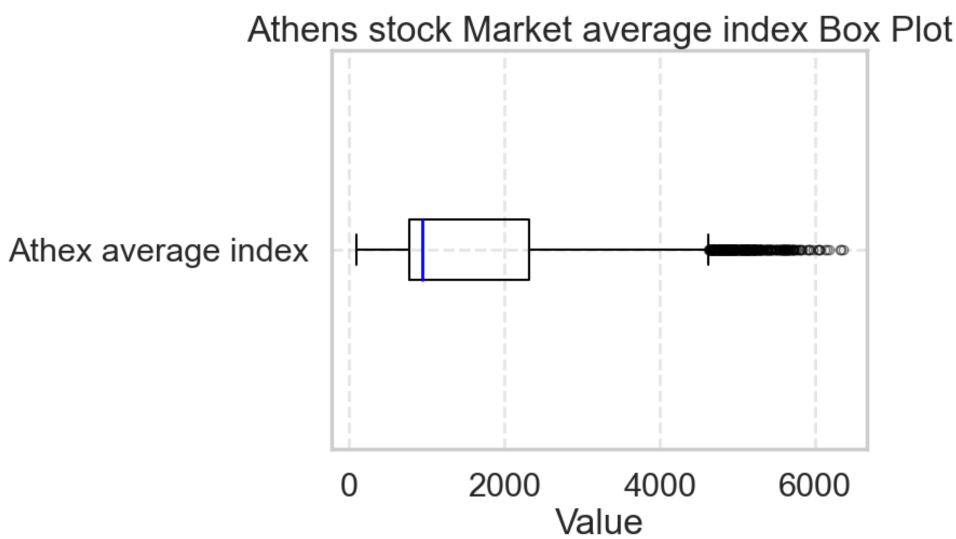
At this point it is important to highlight that although the pearson correlation coefficient is capable of measuring linear relationships amongst variables, it cannot capture any non-linear connections. This dataset has no outliers as it has been prepared and cleaned by the World Bank.This is also empirically confirmed, as all the min and max values of the variables in this dataset are within reasonable limits.

In terms of import prices, it was found in section 3.2- storytelling with data, that this index might be a significant predictor of inflation and hence it constitutes a valuable variable in this research project. Import price data was collected and prepared by the Hellenic statistical authority and hence the quality of the data is expected to be really high, with no missing values, outliers or incorrectly registered values. This is empirically confirmed by generating some basic descriptive statistics. The count of the data is 659, dating back to Januray 1968, there are no missing values and the minimum and maximum values are within reasonable limits. The dataset is quite rich, relatively to other macroeconomic datasets.

CPI and asset prices

The Athens stock exchange general index historic data have been provided by the Bank of Greece, and hence the quality of the data is considered high. However according to section 3.2- storytelling with data, although there is some kind of relationship between CPI and the stock market general index, this relationship is not strong. This is further validated by the Pearson's correlation coefficient of CPI and the average index that only takes a value of 0.16. Additionally, the range of the average index values is 8781, as it has been measured daily every month (weekends and bank holidays excluded) since January the 2nd 1987, and hence the dataset is rich compared to other macroeconomic variables. It should be highlighted that most of the datasets include monthly data and hence this dataset should be prepared with caution before included in a model, in order to ensure the timeliness of the data.

As observed in boxplot 1, bellow, there appear to be many outliers. However, it should be noted that if one sees line-chart 10 (section 3.2-storytelling with data) will realize that the outliers occurred during two specific periods of stock market index spikes, most probably attributed to financial crises. Hence, those will not be considered outliers.



Boxplot 1 – Athens Stock Market avg index. Designed on python, based on ELSTAT data.

In terms of Bond prices and yields, the data quality is high as the dataset is retrieved by the Central Bank of Greece, and hence data cleaning and preparation has been done by the organization (no incorrectly registered values or outliers). However, after exploring the dataset, many missing values are visible for three, five, seven, twenty and thirty year bond prices and yields and hence those variables have been dropped from the excel file.

Ten- and fifteen-year bond prices and yields include all records that date back to January 2002. Although the dataset is not extensive, the relationship between bond prices and inflation as presented in section 3.2- storytelling with data, is clear, and hence the variable may be a strong inflation predictor. The correlation coefficient between CPI and 10 -year price is -0.47, between CPI and 10-year yield 0.30, between CPI and 15-year price is -0.58 and between CPI and 15-year yield is 0.30. Because all bond prices and yields are very highly correlated (Pearson's correlation coefficient of around 0.9), to avoid issues with multicollinearity, only the 15-year price will be considered a potential inflation predictor, as it has the strongest correlation coefficient amongst all (-0.58) and hence the rest of the variables will also be erased from the dataset.

Money Quantity

Money quantity and GDP data have been retrieved from the Hellenic statistical authority (ELSTAT), and hence the quality of the datasets is considered high, as there are no incorrectly registered data, missing values, or outliers. In measuring money quantity, M3 is used, as it is the broadest measure of money supply. It includes all liquid forms of money, along with some longer-term less liquid financial assets such as institutional money market funds (Mishkin, F. S., 2007). The M3 data date back to January 1980 and hence the count of values in this dataset is 515. The dataset has no null values and the Grubbs' test confirms that there are no detected outliers (no maximum z-score is greater than the critical value). In general, due to the high Pearson's correlation coefficient (0.95) between CPI and M3, M3 seems like an important inflation predictor. Finally, it is important to mention that the original dataset also included M1 and M2, however, due to the fact that M3 incorporates both M1 and M3 in its measurements and at the same time values for M1 and M2 are only available from January 2001 onwards, those 2 variables were deleted from the excel sheet and only M3 can be used in the modeling phase.

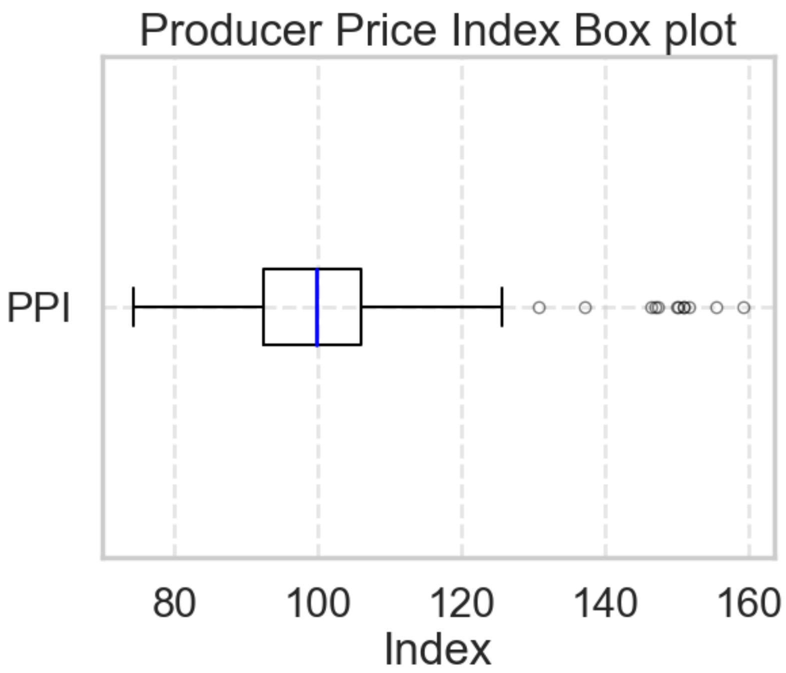
Moving on, as suggested by the quantity theory of money (see section 2- literature review), GDP along with money quantity and money velocity, plays a major role in calculating inflation and hence data collected by the Hellenic statistical authority (ELSTAT) that measure GDP, has been retrieved. However, even though the data has been cleaned and prepared by ELSTAT so that it does not include outliers or incorrectly recorded values, GDP is measured quarterly and hence, in order to ensure timeliness of the data, monthly missing values have been forward filled. In general, forward filling quarterly data is a common technique in macroeconomic analysis, however, it should be noted that the integrity of the data may drop significantly. In any case, the Pearson's correlation coefficient is quite high for the transformed monthly GDP and CPI data (approximately 0.71) and hence GDP is considered a potential inflation predictor, although the dataset's range, even after the forward fill transformation, only includes 327 values.

Industry

In measuring industrial activity, several datasets have been retrieved from the Hellenic statistical authority, namely, business confidence, capacity utilization, industrial production, manufacturing production, and the Producer Price index (PPI). The quality of the data has been ensured by the Hellenic statistical authority and hence the data includes no null values, outliers, or incorrectly registered values. Business confidence has been measured since January 1995 and it includes 457 values. There are no null values, and the minimum and maximum values are within reasonable boundaries (79 and 120) and hence it is confirmed that no outliers are in the dataset. Capacity utilization has been measured since January 1981 and it includes 504 values. The minimum value is 61.3 while the maximum value is 83.8, and hence, it is also confirmed that there are no outliers. Industrial production on the other hand, only dates back to January 2000 and therefore the quantity of the available data shrinks, in comparison to the other indices, as it only includes 275 data points. Manufacturing production also has fewer available measurements in comparison to Capacity utilization and business confidence and ranges back to January 1996, with a total of 323 unique values.

Because the Pearson's correlation between industrial production and manufacturing production is high (the coefficient is over 0.9), only manufacturing production will constitute a potential inflation predictor and be used in future models in order to avoid multicollinearity issues, as it has a larger count of values. Capacity utilization and business confidence also seem to be related as they have a Pearson's correlation coefficient of approximately 0.78. However, this is not too large and hence both variables are considered potential candidates for inflation forecasting.

Moving on, as illustrated in section 3 – Storytelling with data, PPI seems like a significant inflation predictor with a Pearson's correlation coefficient of approximately 0.87 between CPI and PPI. Additionally, PPI data has been collected and prepared by the Hellenic statistical authority (ELSTAT) and is deemed of high quality. However, the quantity of data is quite limited with only 228 data points ranging back to January 2004. Overall, PPI's maximum value is 159, its minimum value is 74.2, while PPI's mean value is approximately 100. Only 25% of the values are bellow 92.3. There are no missing values in the dataset and the line chart 12 presented in section 3 – Story telling with data does not show any signs of existing outliers. This is also confirmed by Box plot 2 that suggests that most values are within the higher and lower adjacent values. It is eminent that some values are much higher than the higher adjacent value and this occurs (according to line chart 2) during the inflation spike starting after 2020, that may have occurred due to COVID19 pandemic and the Energy crisis. PPI will therefore be used as a potential inflation predictor.



Boxplot 2 – Producer price index. Designed on python, based on ELSTAT data.

National Economy

Using measurements that showcase the trajectory of Greece's national economy is imperative in forecasting inflation, as some of those measurements can constitute valuable inflation predictors. To begin with, unemployment rate is usually related to inflation. However, the empirical examination of unemployment data in comparison to CPI discloses that there is no eminent relationship between the two (see appendix 2) and on top of that the quantity of the data is limited as it only dates back to 2004 and is measured in quarterly intervals. Hence, unemployment will not be used in this research project as a potential inflation predictor.

On the other hand, the wage index has proven to be a potentially valuable inflation predictor. First of all, the quality of the data is high since it has been measured and prepared by the Hellenic statistical authority. However, the first record of the index dates back to March 2000 and the measurement has only been recorded quarterly. In order to fill in the null values, the forward fill method has been employed, that fills in the null values with the last recorded value. In section 3.2-storytelling with data, it was found that there appears to be a strong relationship between the wage and CPI. Line chart 15, indicates that the measurement includes no outliers, as its values fall within reasonable boundaries.

Finally, the exploratory data analysis suggests that average interest rates are closely linked to CPI. Although the quality of the data is high, granted that the dataset has been retrieved by the Central Bank of Greece database, the quantity is limited as it only dates to January 2004 (273 data values). There are no missing values nor outliers, as the minimum interest rate is 3.71 and the maximum is 7.25.

It should also be highlighted that EDA suggested that Capital flows, total amount of loans and total amount of deposits, do not show any relationship with CPI and hence those variables will not be considered as potential candidates for inflation prediction in the modeling phase of this report (see appendix 2).

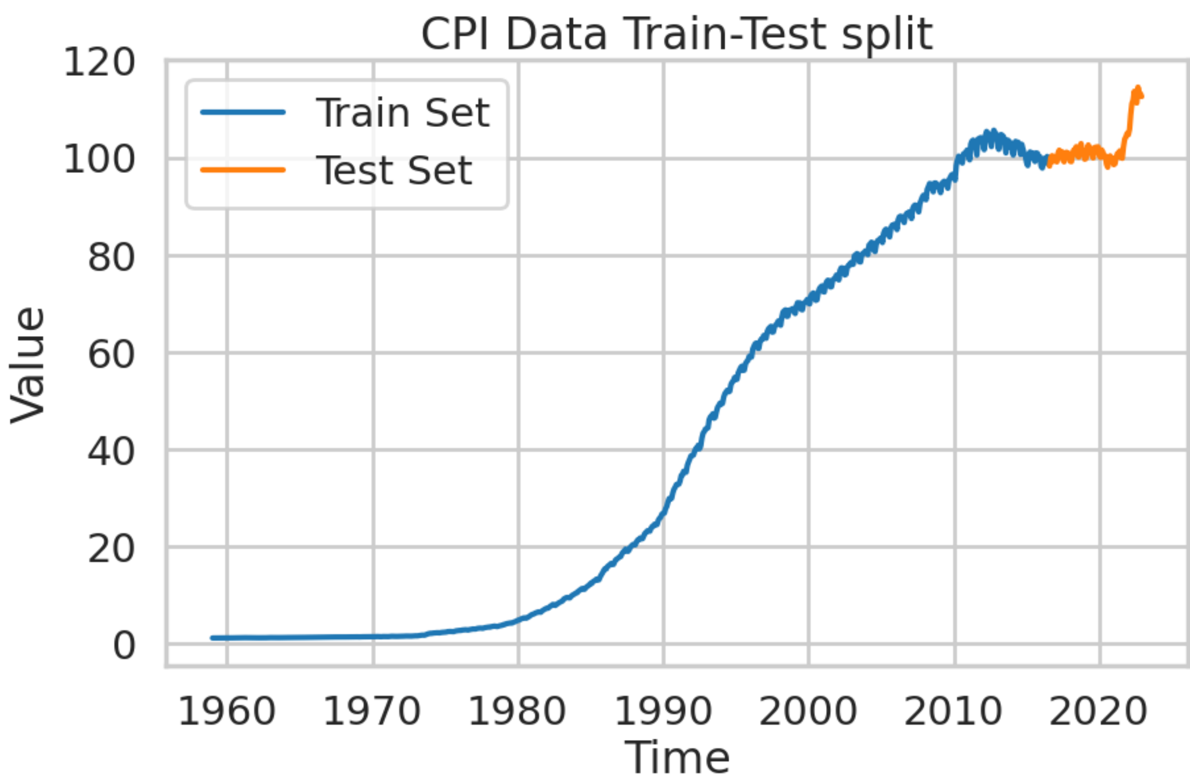
3.4 Modelling & Evaluation

The aim of this report is to compare an ensemble method, specifically an XGBoost model with a deep learning approach, represented by an LSTM ANN, and an ARIMA model, a traditional approach for inflation forecasting. It is hypothesized that an XGBoost model may outperform both traditional and advanced deep learning methods in predicting inflation. The modelling phase of this project begins by designing a persistence baseline which constitutes a performance benchmark for all the proposed methods. Subsequently, the stationarity of the CPI data is ensured, so that an ARIMA model of the right order can be designed. Following, the LSTM architecture is determined, according to theoretical and statistical indications. Finally, the XGBoost model is also developed using theoretical and statistical guidelines. The performance of each model is compared against the other models as well as the baseline model, and the findings are discussed in the final sections of this report.

3.4.1 Train-Test split

First, it is essential that the data are split into a train and test set so that the integrity of the predictive ability of each model is ensured. The train set is utilized to train the algorithms using the available data, while the test set serves the purpose of evaluating the performance of the algorithms on unseen, out-of-sample data. In this manner, it is ensured that information about the data does not leak from the test-set to the training process, and hence, when a model is brought into production, its predictive ability is not compromised. It is also important to mention that a validation set might be useful when designing advanced machine learning models, which is essentially a sub-set of the labeled data used to optimize the model's hyperparameters and architecture. Cross-validation can also be useful in improving the performance of the models for reasons of performance estimation reliability and data scarcity mitigation.

In splitting the CPI data in a train and test set, the data values are not shuffled, and the order of the data is maintained, as this is an integral requirement in time-series forecasting. Due to the small range of CPI values, only 10% of the values are assigned to the test set, while 90% of the values are assigned to the train set (and potentially the last 10% of those to a validation set), as illustrated in line chart 16.



Line chart 16 - CPI data Train-Test split. Designed on python, based on ELSTAT data.

3.4.2 Measuring performance

In measuring the performance of the models, the Root-Mean-Squared-Error (RMSE) is calculated. Then the RMSE score of all models will be compared to conclude which model has the highest predictive ability. Additionally, RMSE will be used for training the models and when adjusting the hyperparameters of the models. It is a commonly used metric in evaluating the performance of various models in machine learning, and it measures the magnitude of the differences between a model's predicted values and the actual observed values. The smaller the RMSE value, the better the model is performing, indicating that the model's prediction if closer to the actual observed values.

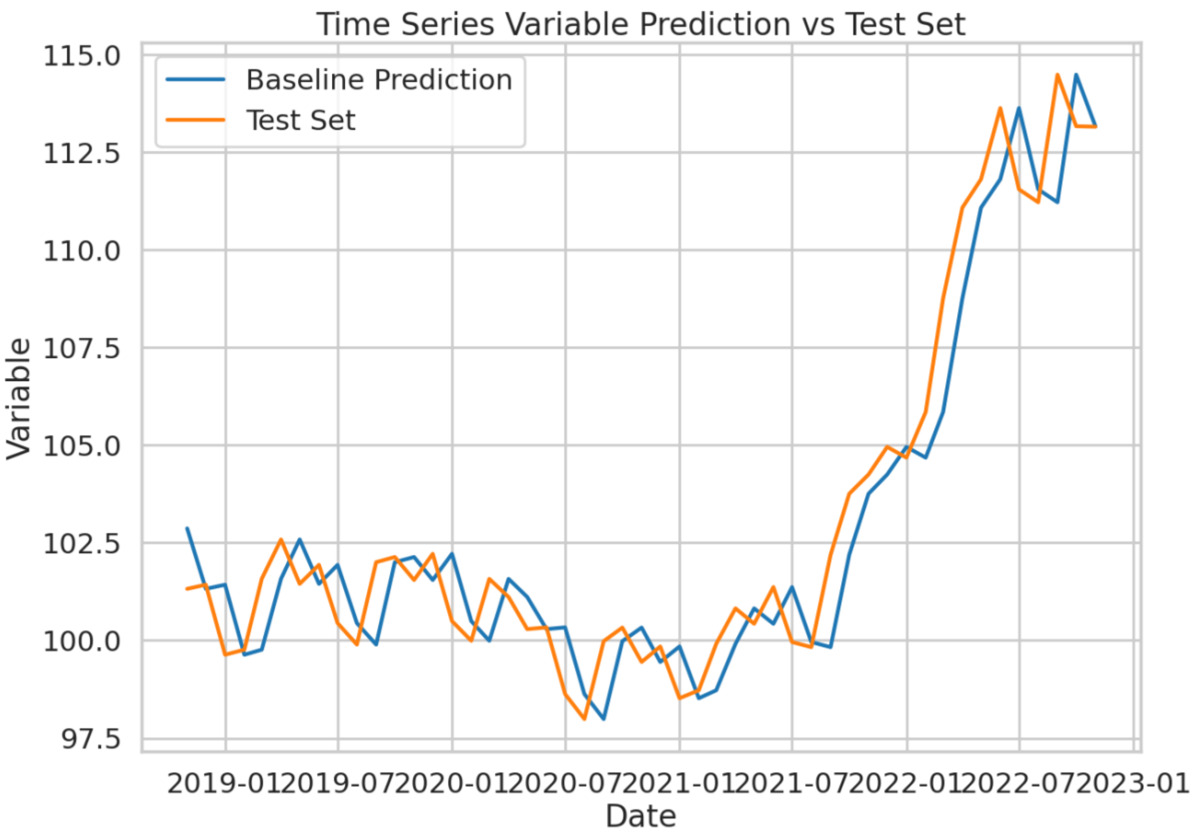
Mathematically, RMSE is calculated by taking the root of the mean of the squared differences between the predicted CPI values and the actual CPI values as such:

RMSE = $\sqrt{mean((predicted - actual)^2)}$

3.4.3 Persistence baseline

In the context of data analysis, a baseline refers to a simple, basic, and rather intuitive model that establishes a benchmark for evaluating and comparing the performance of more complex algorithms with. The persistence baseline is implemented by forecasting future values of the target variable as identical to the most recent observed value, resulting in a simple horizontal shift of the target variable. A persistence model is deemed appropriate to be used as a performance benchmark in this project, as it is simple to build, requires minimal computational power and it is transparent and easy to interpret. Additionally, a persistence benchmark in inflation forecasting, is quite challenging and hard to surpass, as inflation is considered a relatively persistent measurement, especially in short time horizons. Hence, a persistence baseline model is considered a relevant and challenging baseline in the context of this project. However, it should be mentioned, that a persistence model should not be considered a reliable and trustworthy model, as it completely fails to capture the directional changes in CPI values and hence its predictive ability is compromised.

In the case of this research project, because the time-horizon for inflation forecasting is just 1 month, the CPI values are shifted one step forward. The forecasted CPI in time t is hence the actual CPI value in time t-1. It is eminent that because all values are shifted one month ahead, the first value in the test set is empty and hence the row is deleted. It should be highlighted, that there is no need for a validation set in designing the persistence baseline, as this is a very simple model that does not require any kind of hyperparameter tuning. The RMSE of the persistence baseline is approximately 1.314.



Line chart 17 – persistence baseline CPI predictions and actual CPI values. Designed on python.

3.4.4 ARIMA

Testing for stationarity

It is essential, in order to identify the order of the ARIMA model to make the time-series data stationary. In general, the ARIMA model inherently assumes that the time series data is stationary, meaning that its statistical properties(mean and variance) do not change over time. If this underlying assumption does not hold, the resulting forecasts may be unreliable. Essentially, making the time series data stationary, invlolves removing the trends or patterns and adresses the presense of seasonality in the data, that make the series non-stationary.

The first approach in assesing the stationarity of the series, is to visually identify whether there is any sort of trend or seasonality present in the dataset. It is eminent that CPI has a clear upward trend and some sort of seasonality (see line chart 2). The mean and the variance seem to change over time and hence the series look non-stationary. However, in data analysis, it is always safer to confirm a hypothesis with multiple methods. Hence, additionally to the visual identification of non-stationarity, the ADF test of stationarity and the KPSS test of stationarity are also performed.

The ADF test's null hypothesis is that of non-stationarity(unit root) of the time series data, while the alternative hypothesis is that of stationarity. The results of the ADF test performed on the CPI dataset indicate that the p-value of the ADF statistic (0.50) is approximately 0.98, well above the significance threshold of 0.05 and therefore, the null-hypothesis of non-stationarity cannot be rejected, against the alternative hypothesis of stationarity. Hence, the ADF test confirms the visual suggestion of non-stationarity.

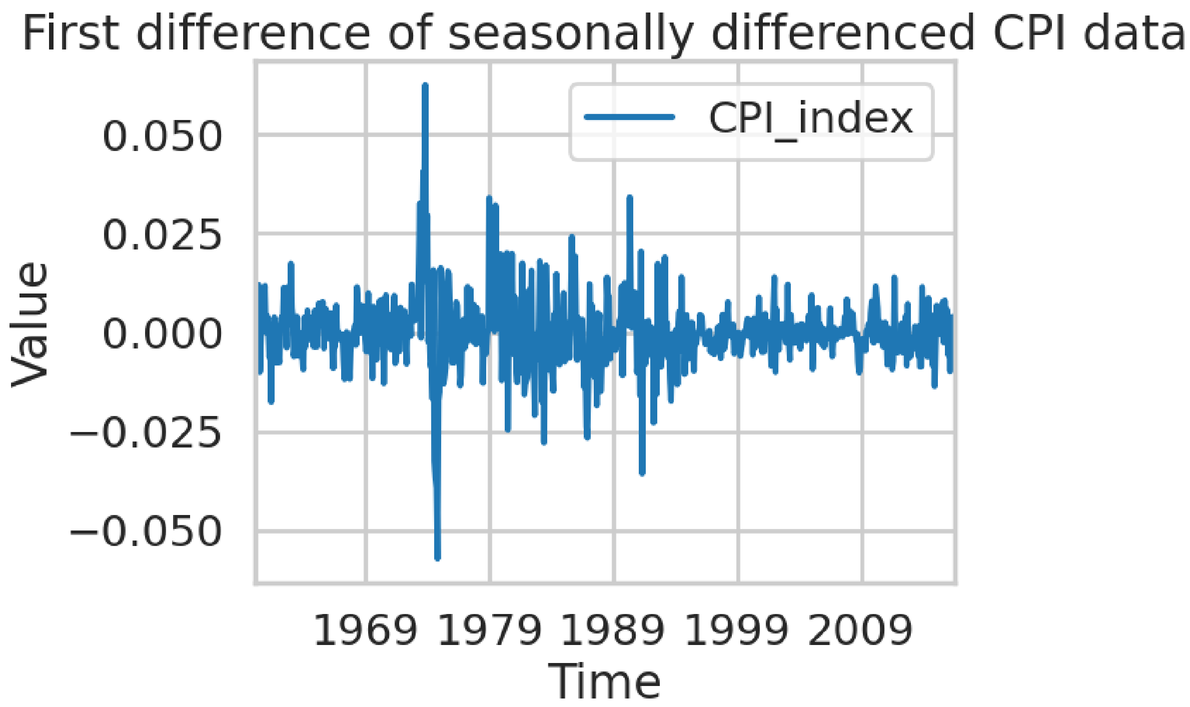
The KPSS test's null hypothesis is that the time series data is stationary, while the alternative hypothesis is that of non-stationarity. The results of the test suggest that the p-value of the KPSS test statistic (3.64) is approximately 0.01, bellow the significance threshold of 0.05 and therefore, the null-hypothesis of stationarity can be rejected and the alternative hypothesis of non-stationarity is accepted. Hence, the KPSS test confirms both the visual and ADF test suggestion of non-stationarity. Therefore, the CPI time-series data is considered non-stationary and an ARIMA model cannot be fitted unless the series is rendered stationary.

Stationarizing the CPI time-series

There are many approaches to make time-series data stationary, such as logarithmic transformation, square root transformation, Box-Cox Transformation, differencing, seasonal differencing, exponential smoothing, and others (Hyndman & Athanasopoulos, 2018, Section 8.4). However, not all approaches are a good fit for every time series, as a successful transformation is subject to each dataset's unique characteristics.

In the case of this research project, various approaches have been tested, most of which did not make the CPI time series stationary. The square root transformation, along with the logarithmic transformation result in a non-stationary time series (see appendix 4). Transforming the data with first degree differencing also results in a non-stationary time-series (see appendix 4). Second degree differencing seem lead to a stationary time-series but the degree of the ARIMA model cannot be identified, as the autocorrelation function plot and the partial autocorrelation function plot only include significant autocorrelation and partial autocorrelation spikes for most lags (see appendix 4). Calculating the seasonal percentage change also does not render the time-series data stationary (see appendix 4).

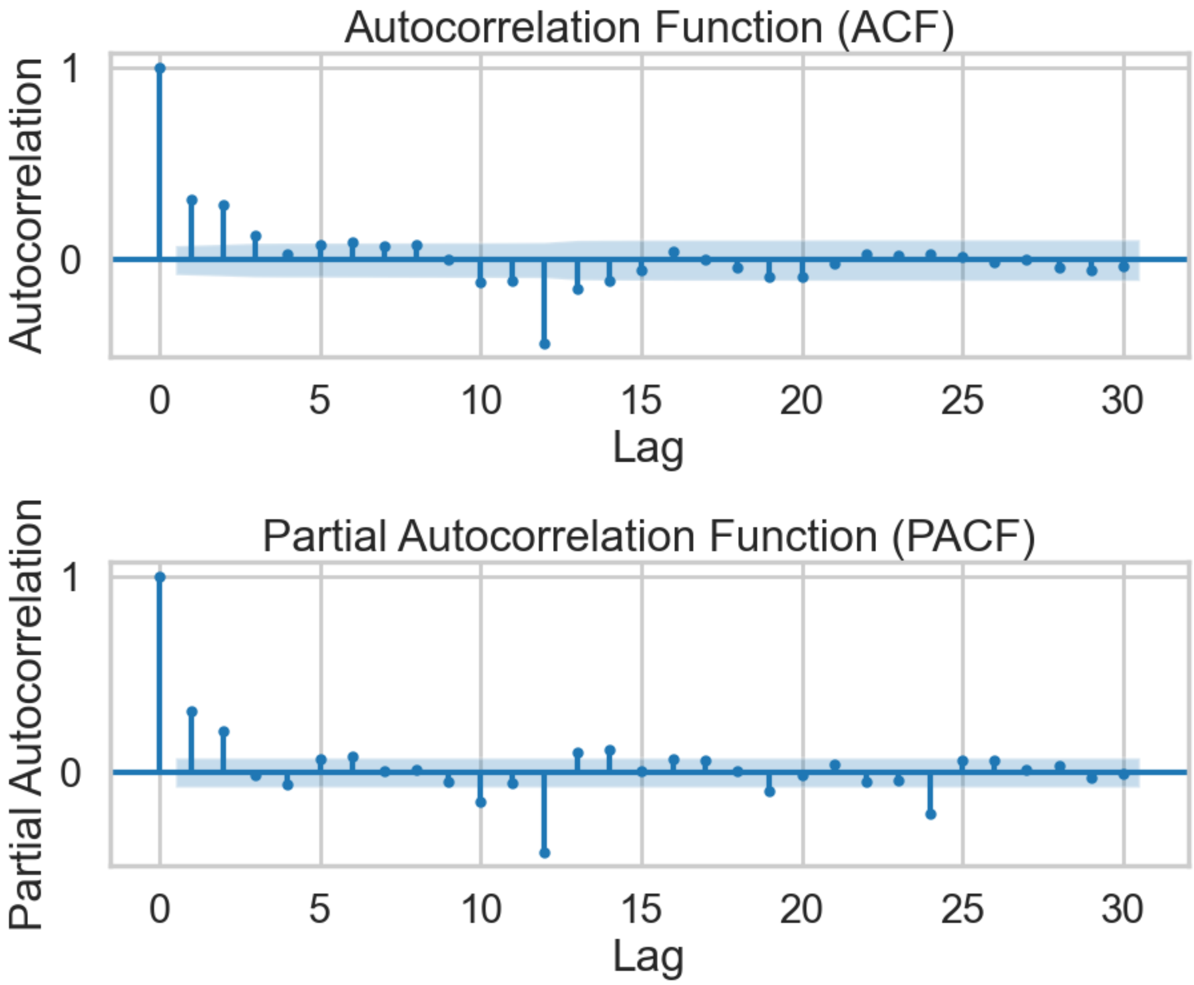
Finally, taking the seasonal percentage change of the values (12-month season), and then taking the first difference of the seasonal percentage change, results in the stationarity of the CPI time-series data. In essence, the percentage change between time t and time t-12 of the CPI values is calculated, by finding the difference between t and t-12 and then divide the difference by t-12 and multiply it by 100. Thereafter, the first difference between the percentage change at time t and t-1, with t > 14, is also calculated. This will be used as input in the ARIMA model. The ADF's test statistic on the first difference of the seasonal percentage change of the CPI data, takes a value very close to 0 and well below the 0.05 significance threshold and hence, the null hypothesis of non-stationarity can be rejected, and the alternative hypothesis of stationarity can be accepted. The KPSS test's p-value is 0.1, above the 0.05 significance threshold and hence the null hypothesis of stationarity cannot be rejected, against the alternative of non-stationarity (unit root). The stationarity of the dataset can also be confirmed by plotting the first difference of the seasonal percentage change of CPI values (see line chart 18). However, it should be highlighted that although the differenced data seem to have a stable mean and variance, as time progresses, the fluctuations around the mean change. However, it is assumed that this does not violate the stationarity assumption as the mean and variance seem to not change significantly over time and hence will not pose any problems in determining the order of the ARIMA nor in generating a capable ARIMA model.



Line chart 18 – First difference of the seasonal percentage change of the CPI values. Designed on python.

Determining the order of the ARIMA model

In determining the order of the autoregressive part (AR) of the ARIMA it is important to look at the ACF plot. The autocorrelations seem to decay gradually, with the first 3 lags being significant and the fourth lag marginally significant. Hence, the order of the AR part of ARIMA is 4. Similarly, the first 3 lags of the Partial autocorrelation function plot, are significant, while the rest (with a few sporadic exceptions) enter the insignificant zone and hence the order of the Moving average (MA) part of ARIMA is also 4. Therefore, the suggested model is ARIMA (4,0,4) since the data have already been differenced.



Bar-chart 1 - Autocorrelation and partial autocorrelation plots for ARIMA(4,0,4). Designed on python.

Fitting the ARIMA(4,0,4) model

After the order of the ARIMA has been determined, the model must be fitted on the training data. The summary of the model shows that all lags of both the AR and MA part of the ARIMA are significant, and the coefficients have been determined (see table 5).

	coef	std err	z	P>|z|	[0.025	0.975]
const	-4.226e-05	0.001	-0.074	0.941	-0.001	0.001
ar.L1	1.6209	0.215	7.537	0.000	1.199	2.042
ar.L2	-1.5162	0.210	-7.232	0.000	-1.927	-1.105
ar.L3	1.0437	0.190	5.481	0.000	0.670	1.417
ar.L4	-0.3844	0.124	-3.090	0.002	-0.628	-0.141
ma.L1	-1.4176	0.213	-6.651	0.000	-1.835	-1.000
ma.L2	1.4969	0.159	9.413	0.000	1.185	1.809
ma.L3	-1.1811	0.201	-5.867	0.000	-1.576	-0.787
ma.L4	0.5056	0.122	4.157	0.000	0.267	0.744
sigma2	6.405e-05	2.3e-06	27.806	0.000	5.95e-05	6.86e-05

Table 5 – ARIMA(4,0,4) summary table.

Now the test data can also be made stationary, as per the train data indications, to avoid data leakage. Hence, the test data are seasonally differenced in terms of percentage and then the first differences are calculated.

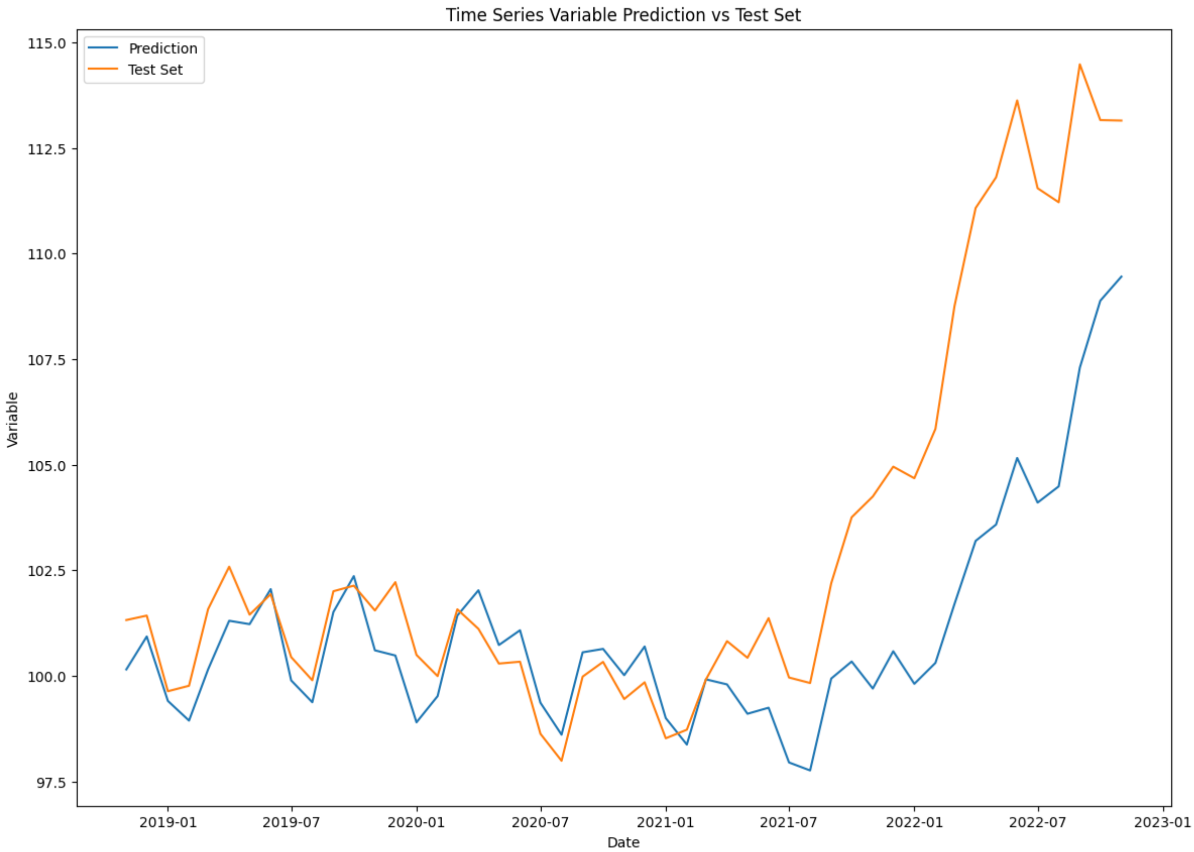
Testing ARIMA (4,0,4)

In testing the performance of ARIMA(4,0,4), the model has been fitted on the train set and forecasts the CPI values for the next month (one step ahead) after considering the autocorrelations and partial-autocorrelations for the previous four months. This process iterates in the test set until the model has forecasted all the dates present in the set. Since this is an one step ahead forecast, after each forecast is made, the list of inputs is modified to include the actual test-set observations.

Because the first difference of the seasonal percentage change of the original data is calculated to predict the CPI, the first fourteen values of the test set are not included in the prediction series (12 values pertaining to the seasonal differencing + 1 value pertaining to the first difference + 1 value pertaining to the 1-month time horizon). It should be mentioned that for reasons of comparability, those fourteen values are also taken out of both the LSTM and XGBoost models' test sets. In general, the performance of all models is assessed on the same unseen data to ensure that their respective performance is comparable.

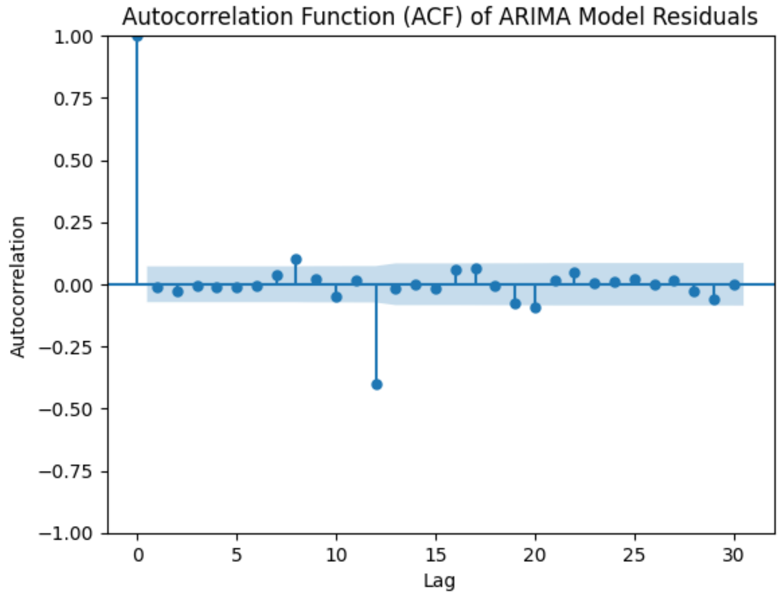
The cumulative sum of the predictions (after the 14^{th} prediction) is then calculated (the reversed seasonal percentage change data). The test set's actual observations are shifted 12 months forward, and starting after the 14^{th} observation, the actual test set values are multiplied by 1 + the reversed seasonal percentage change data to get the forecasted CPI values.

The RMSE of the model is 3.440, quite higher than the persistence baseline's RMSE. However, this is quite usual, since CPI is quite a persistent measurement, and the time horizon of just 1 month is narrow for significant changes in the CPI values to occur. It is interesting to see in line chart 19 that until 2022 the forecasted values are quite similar to the actual observations and thereafter, perhaps due to the COVID19 pandemic and the Russian invasion of Ukraine, the predicted values fall behind in comparison to the actual observations. In general, it is very hard for an ARIMA model to capture such unpredictable turning points.

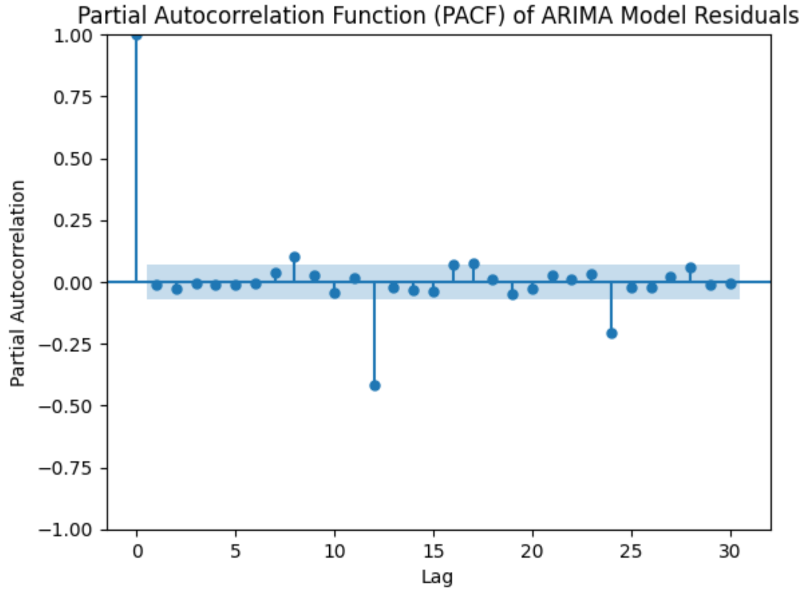


Line chart 19 – ARIMA(4,0,4) CPI forecasts on the test set and CPI actual observations. Designed on python.

As outlined in the ARIMA methodology process in section 2- literature review, residual diagnostic checks are detrimental in assessing the model's adequacy. Hence, the ACF and PACF plots of the residuals are generated in order to assess whether the model is adequate and there are no temporal dependencies or autocorrelation in the residuals. The ACF and PACF plots of the residuals suggest that lag 12 of the autocorrelation plot and lag 12 and 24 of the partial-autocorrelation plot exceed the significance threshold, signifying that there still may be some temporal dependency or autocorrelation in the residuals. This is expected, as the CPI autocorrelation and partial-autocorrelation plots also suggest significance for those lags. Hence, it is indicated that the model may not adequately capture all the underlying patterns or information in the data. However, an ARIMA model is fundamentally not capable of capturing complex relationships and turning points as it is essentially agnostic and backward looking, and hence, alternative models need to be explored in the domain of inflation forecasting.



Bar chart 1 – ARIMA(4,0,4) ACF residuals.



Bar chart 2 – ARIMA(4,0,4) PACF residuals.

3.4.5 LSTM Artificial Neural Networks

Variable selction for multivariate LSTM ANN

In designing an optimal LSTM ANN, it is essential that the most relevant variables for CPI forecasting are selected. The selection process involves considering the variable size and its relationship with the CPI, as outlined in section 3.2 - Storytelling with data. It is important to note that due to the limited amount of available CPI data, a large number of predictors may lead to model overfitting and low performance, and hence, it is necessary to exercise caution when selecting the model's input variables. Hence from each input category, the most appropriate variables as per the EDA are selected. Additionally, considering the observed seasonality in CPI on an annual basis, the lagged values of each input variable for the twelve-month period preceding the forecasted month are included in the model. This approach is based on the notion that utilizing data from the past twelve months can provide valuable insights and capture significant relationships for accurate forecasting.

In determining future CPI values, CPI's lagged values should be included in the model's inputs. In terms of energy, considering that Crude oil Average Price and Natural Gas EU Price, are very significant variables for inflation forecasting according to the EDA, their lagged values for the past twelve months will also be included in the model's input list. Furthermore, in terms of Money Quantity, although M3 decreases the size of available data, as the first recorded M3 value only dates back to 1980, the EDA showed that M3 may also be one of the most significant predictors of CPI. Hence, M3 lagged values will also be included in the model's input list. All asset price variables include only a very limited amount of data and hence because this may negatively affect the performance of the model, none of the variables will be included in the input list. The same occurs with most national economy indices except of import prices. The exploratory data analysis suggests that import prices may be a strong determinant of CPI. Additionally, import prices date back to 1968 and hence, this variable will not decrease the data size. Therefore, import prices will also be used as an input in the LSTM. Finally, in terms of Industrial indices, only the inclusion of capacity utilization is deemed appropriate, as it was found that it is strongly related to CPI and additionally, the range of the variable's values is sufficient and does not affect the timeliness of the data. Therefore, the LSTM ANN model will be built using twelve lagged values of the following variables: CPI, M3, import prices, Crude oil average price, Natural Gas EU, and Capacity utilization. The available data, ranges from January 1981 to November 2022.

Data preprocessing

The LSTM ANN is designed using Python’ s Pytorch library. Therefore, the shape of the inputs and target variable is adjusted so that they fit Pytorch’s requirements. All the data are transformed to be NumPy arrays of shape: “batch size, sequence length, number of features”. Batch size refers to the number of samples in each batch of data, sequence length refers to the length of the input sequence (lookback) and the number of features refers to the number of input features. In deciding the batch size, 2 main factors are taken into consideration: Computational efficiency and Generalization. The larger the batch size, the faster the training process (Loshchilov & Hutter, 2018) and the lower the generalization capability, and the smaller the batch size, the slower the training process and the higher the generalization capability (Keskar et al., 2016). Because the dataset used in this research project is relatively small, a large batch size would reduce generalization by much and hence a small or moderate batch size should be used. After designing the model architecture, and during hyperparameter tuning, experimentation with batch sizes of 32 and 64 showed that the best batch size for the selected model architecture is 64. In general, the shape of the features is (64, 12, 6). The shape of the target is (64, 12, 1).

The data is split into a train set, a validation set and a test set. The train set consists of 80% of the data, the validation set of 10 % of the data and the test set also consists of 10% of the data. It is important to note that the order of the train, validation and test set should remain intact. When dealing with time-series analysis, shuffling the order of the sets will most probably lead to the complete inability of the model to perform. The purpose of the train set is to train the model, the validation set instrumentalizes the hyperparameter tuning process, while the out-of-sample test set is used to assess the model’s performance on unseen data.

In preparing the data for the LSTM ANN it is vital, that the data are first standardized. In this manner, if input variables are of different scales, they should be transformed and adjusted to the same scale so that the LSTM algorithm can model the data more appropriately. In doing so, the data are transformed using the standard scaler. In essence, the standard scaler standardizes a feature by subtracting the mean and then scaling to unit variance, meaning dividing all the values by the standard deviation. The scaler is fitted only on the train set in order to avoid data leakage, meaning information about the test set leaking into the training process. It should be noted that all variables are finally saved as Pytorch tensors before being forwarded to the LSTM ANN.

LSTM ANN Architecture

In designing the LSTM Artificial Neural Network, different combinations of hyperparameters are used iteratively on different model designs, starting from the simplest model architecture (with less layers and no activation functions) and gradually increasing model complexity (see appendix 5).

The best model is found to have 1 LSTM layer, 3 fully connected layers (fc1, fc2 and fc3) and the ReLU activation function. More specifically, the LSTM layer processes the input sequence (12 months lookback for 6 different variables), updates its hidden state, and produces the output sequence. The best combination of hidden state and number of stacked layers is determined during hyperparameter tuning. The output sequence is then forwarded to fc1 that takes the output of LSTM layer as input and maps it to a higher-dimensional space with 64 units. Thereafter, the output of fc1 goes through the ReLU activation function. Thereafter, fc2 takes the output of the ReLU activation function as input and maps it to a lower-dimensional space with 32 units. Finally, fc3 is the final fully connected layer. Because the model is designed for one step ahead time-series forecasting, fc3 is a linear transformation function that takes the output of fc2 ant maps it to a single unit. This unit is the CPI value one step ahead (see diagram 1). Adding less complexity (such as fewer fully connected layers) or more complexity (such as more fully connected layers) to the model architecture, has proven to drop the model's performance (see appendix 5 for other model architectures and their corresponding performance).

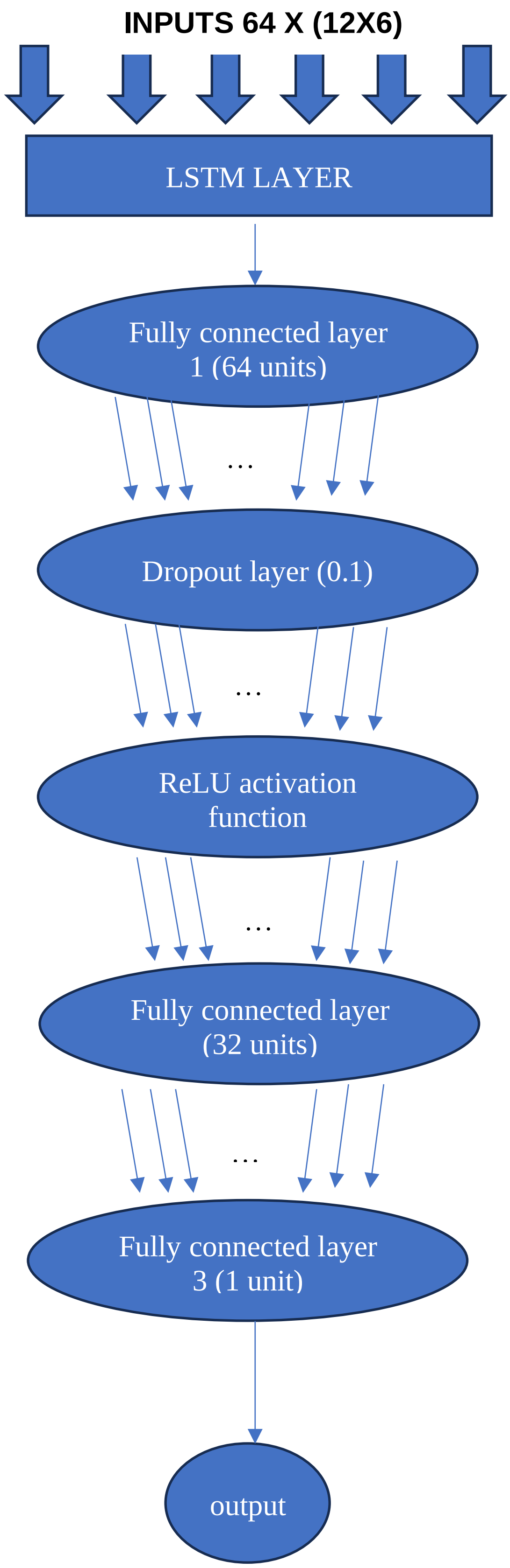


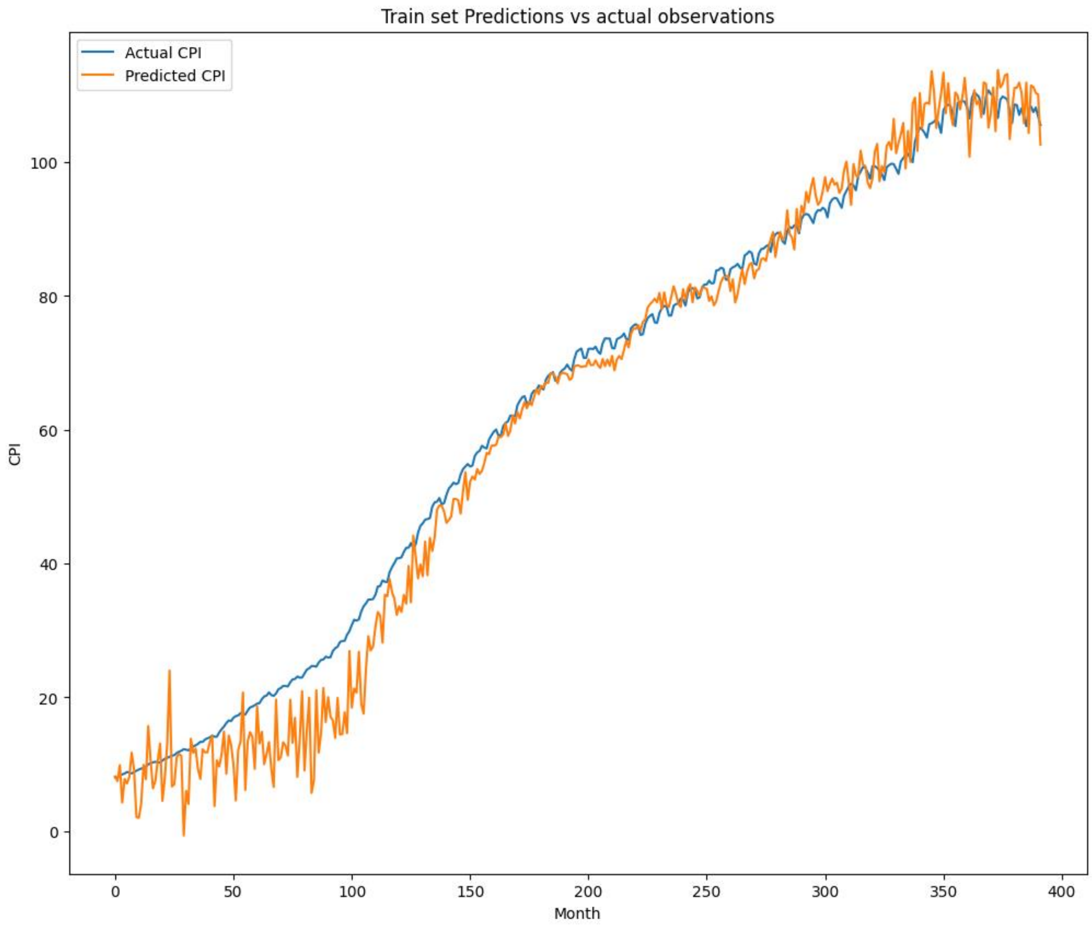
Diagram 1 - LSTM Network Architecture

During training, the Adam optimizer is used for backpropagation. Adam is an extension of the stochastic gradient descent (SGD), that incorporates adaptive learning rates for individual parameters as well as the momentum term, helping accelerate the learning process (Kingma, D, P., Ba, J. , 2014). Using the Ray Tune package in Python, different hyperparameter combinations are used in training (hidden size, number of stacked layers, learning rate, epochs, and batch size) and then each model is evaluated on the evaluation set. The best performing model in the evaluation set is the selected model. It should be mentioned that the hyperparameter space along with the model architecture are adjusted multiple times, to reach better results, according to model performance. The best performing model is that of hidden size of 60, 6 stacked layers, 1000 epochs, a learning rate of 0.001 and batch size of 64, while the Network's architecture is the one described previously (see appendix 5 for models with different architectures, hyperparameter combinations and their respective RMSEs).

Considering the limited availability of data, it is important to acknowledge that all models exhibit some degree of overfitting. To mitigate this issue, the dropout regularization technique is applied. In this context, the LSTM model incorporates a dropout layer with a dropout rate of 0.1. This dropout layer is specifically added after the fc1 layer. The purpose of dropout is to randomly set a fraction of input units to zero during each training iteration, reducing interdependencies among units and encouraging the model to learn more generalized representations (Goodfellow, I., et al., 2016; see appendix 5 for models with no regularization).

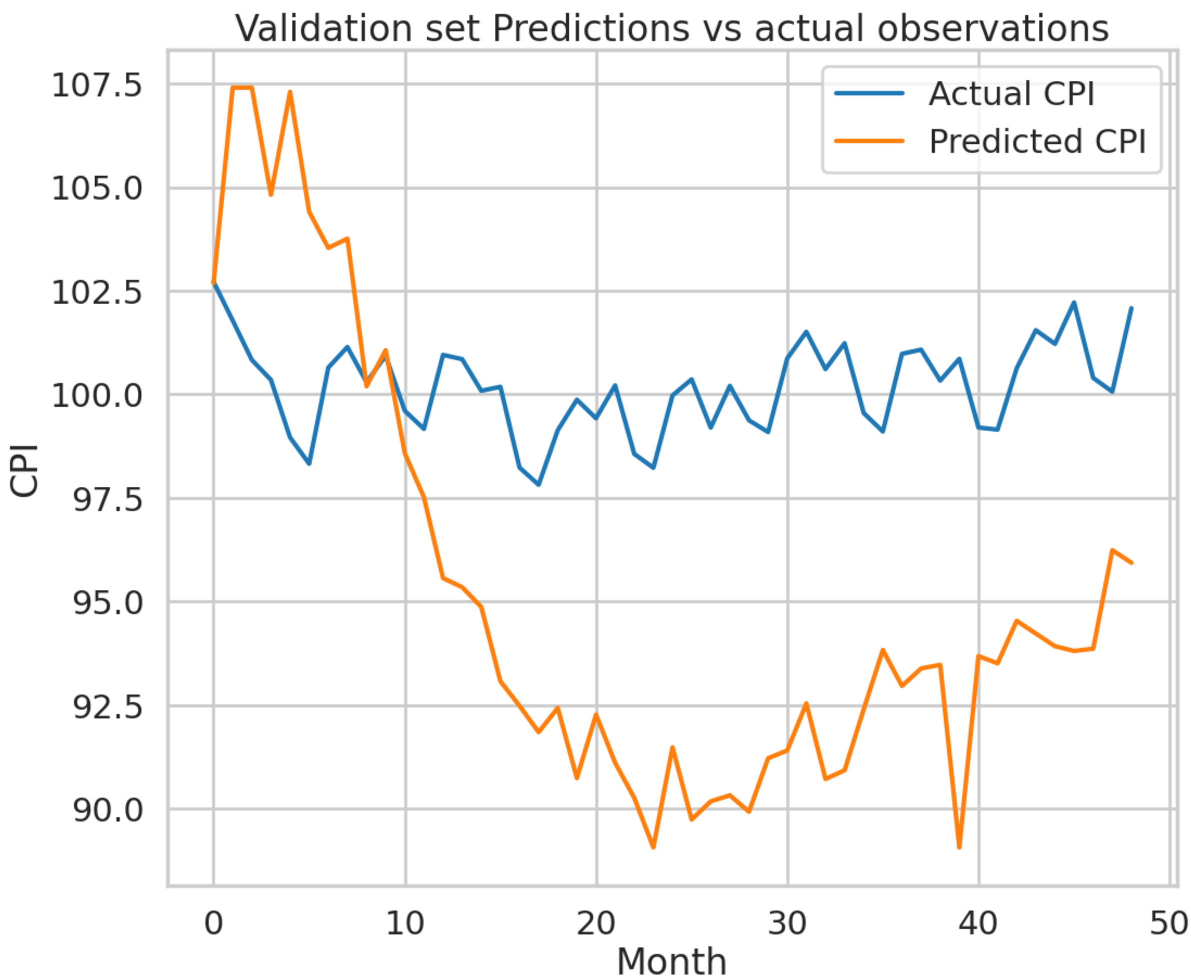
LSTM ANN results

To begin with, it is important to highlight that because all data were scaled before training the model, the scaling process is reversed to illustrate the actual predicted CPI values. In line chart 20, the train set predictions against the actual observations are illustrated in the same plot. It is eminent that the predictions line follows the direction of the CPI actual observations trend line, however, the fluctuations are rather intense. This occurs due to the dropout layer included in the model architecture aiming at decreasing overfitting. It is visible, however, that the LSTM ANN model does not perform optimally in predicting one step ahead CPI values.



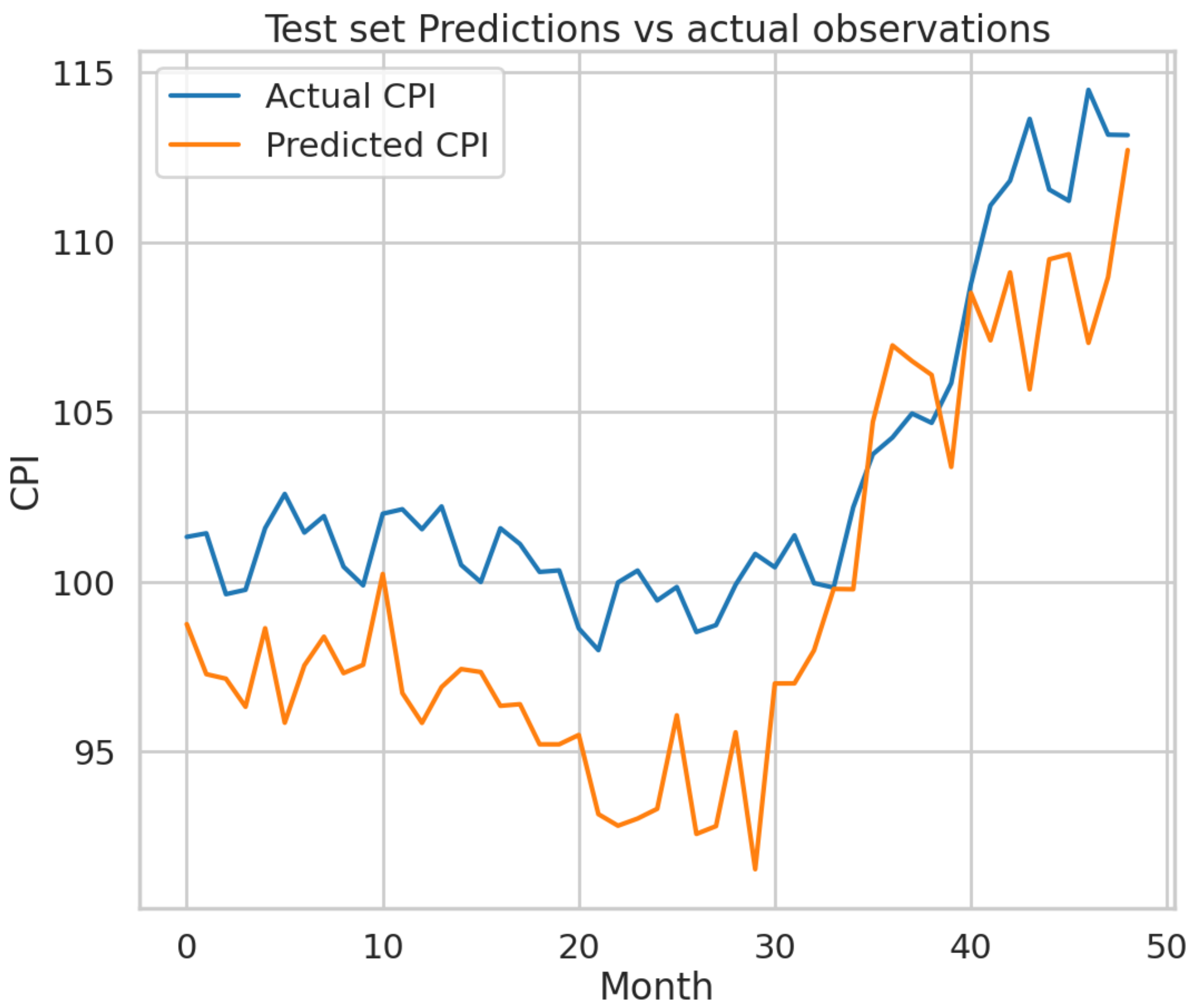
Line chart 20 – LSTM CPI forecasts on the train set and CPI actual observations. Designed on python.

During validation, overfitting becomes more apparent. Although the model's predicted line closely tracks the direction of the actual CPI values trend line, which is partly attributed to the inclusion of the dropout layer, the model consistently underestimates the CPI values. This behavior may be influenced by the distribution of actual CPI values, which exhibits a positive skewness, as it is illustrated in appendix 6, line chart 39 and in CPI's summary statistics. This indicates that the dataset contains a higher proportion of smaller values compared to larger values. Consequently, the model "learns" that its overall performance is improved when the predictions are lower rather than higher. However, in time-series forecasting, where the validation set follows the training set chronologically, it is expected that the actual observations in the validation set will generally be higher than a significant portion of the training set. As a result, the model's forecasts tend to underestimate the actual observations.



Line chart 21 – LSTM CPI forecasts on the validation set and CPI actual observations. Designed on python.

The suggestion for CPI values underestimation is further validated if one looks at the test set in line chart 22, where although the direction of the forecasted CPI values seems to be quite accurate, the magnitude of the predicted values is constantly underestimated throughout the test set. It should be highlighted that after 2022, the CPI values recorded an all-time high, with an unprecedented inflation spike (starting after the 35th observation in the test set; see line chart 22) that rises to a CPI of around 115. This has most likely occurred due to the COVID19 pandemic and the Russian invasion of Ukraine (see section 2 – Literature review). It is important to mention that the LSTM ANN model was able to capture this complex and unprecedented inflation spike to a great extent, as opposed to all the other models employed in this report, highlighting that deep learning models and neural networks, may have an increased ability compared to other models in capturing complex relationships.



Line chart 22 – LSTM CPI forecasts on the test set and CPI actual observations. Designed on python.

The RMSE of the best performing LSTM model on the test set is 3.74, well above the persistence baseline's RMSE (1.314) and slightly higher than the ARIMA (4, 0, 4) model's RMSE (3.440). Of course, it should be once more highlighted that the persistence baseline, although it might have a significantly lower RMSE, since inflation is a persistent macroeconomic measurement, it does not account for the direction of the CPI values changes.

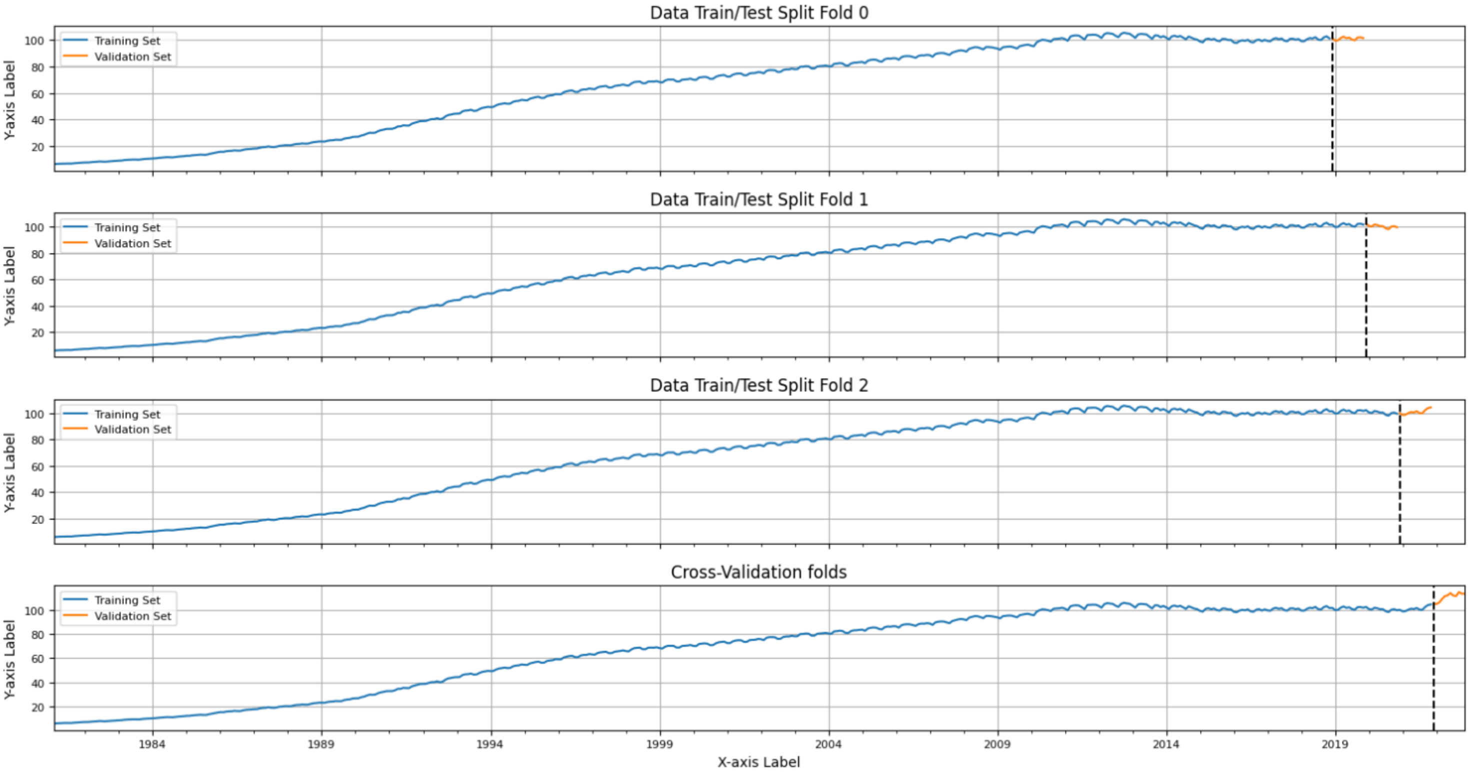
3.4.6 XGBOOST Regressor

Data Preprocessing

The selected variables for the XGBoost are the same as in the LSTM ANN, according to the EDA's suggestions. It is important to mention that when designing an XGBoost regressor for time series forecasting, it is advisable that the features' lookback (the number of lagged values) should not exceed the forecasting horizon, as this may lead to increased model overfitting risk, as the model may become too complex and may memorize noise and idiosyncrasies in the training data rather than learning general patterns. Additionally, a lookback window larger than the forecasting window may lead to delayed response, as the predictions will be based on information that is distant in time from the target variable, resulting in delayed response to changes in the underlying patterns of the time series (Chen, T., et al., 2016). Hence, because the objective of this project is to forecast CPI at time t, each variable will only have one lagged value, at time t-1. Additionally, although it is advisable to normalize the data before entering them to an XGBoost regressor, since Decision trees and more specifically, the XGBoost, is not as sensitive to variations in data scales as the Artificial Neural Networks are, it is deemed that preserving the original data will not pose any issues to the modelling phase.

The package that is used to design the XGBoost regressor is Python's xgboost. Xgboost can accept either a NumPy arrays or a Pandas series, and hence, all the relevant lagged values of the features along with the target variable, CPI at time t, will be stored in a Pandas DataFrame and then the relevant series for the inputs along with the target variable series will be inserted to the XGBoost regressor.

As always, it is imperative to split the data into a train and test set. To increase model performance, 'TimeSeriesSplit' from scikit-learn's 'model_selection' module, a specific type of cross validation strategy designed for time series data, will be used. The strategy is to create 4 consecutive folds, or splits that will be utilized for validating the model after it is trained as illustrated in line chart 23, bellow.



Line chart 23 – Train-test split cross-validation folds. Designed on python.

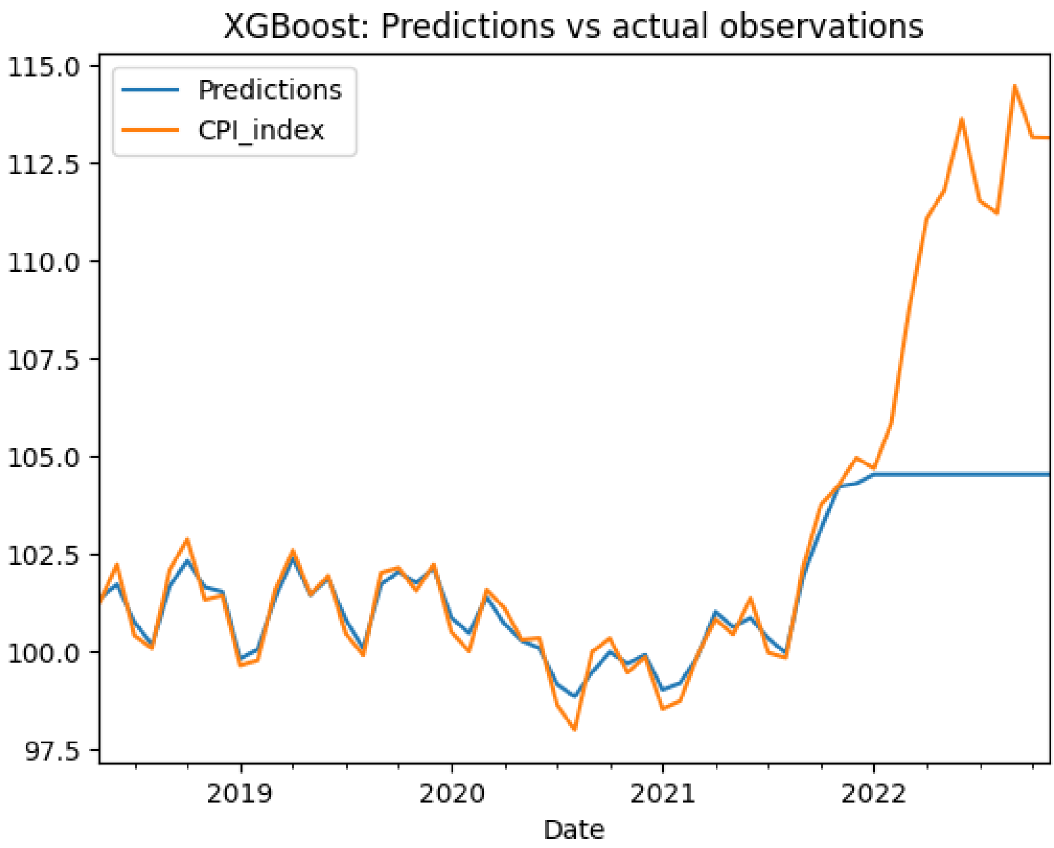
Model design and hyperparameters tuning

In designing the xgboost regressor different number estimators and learning rates were used, starting with a low number of estimators (1000 estimators) and gradually increasing it while combining it with a learning rate of 0.01 and 0.001. The optimal number of estimators was found to be 8000 while the optimal learning rate was found to be that of 0.001. The loss function chosen for the purpose of this project is RMSE (root-mean-squared-error), while the gbtree booster is brought into use. The model is trained on the train set and evaluated on the test sets as per the cross-validation strategy.

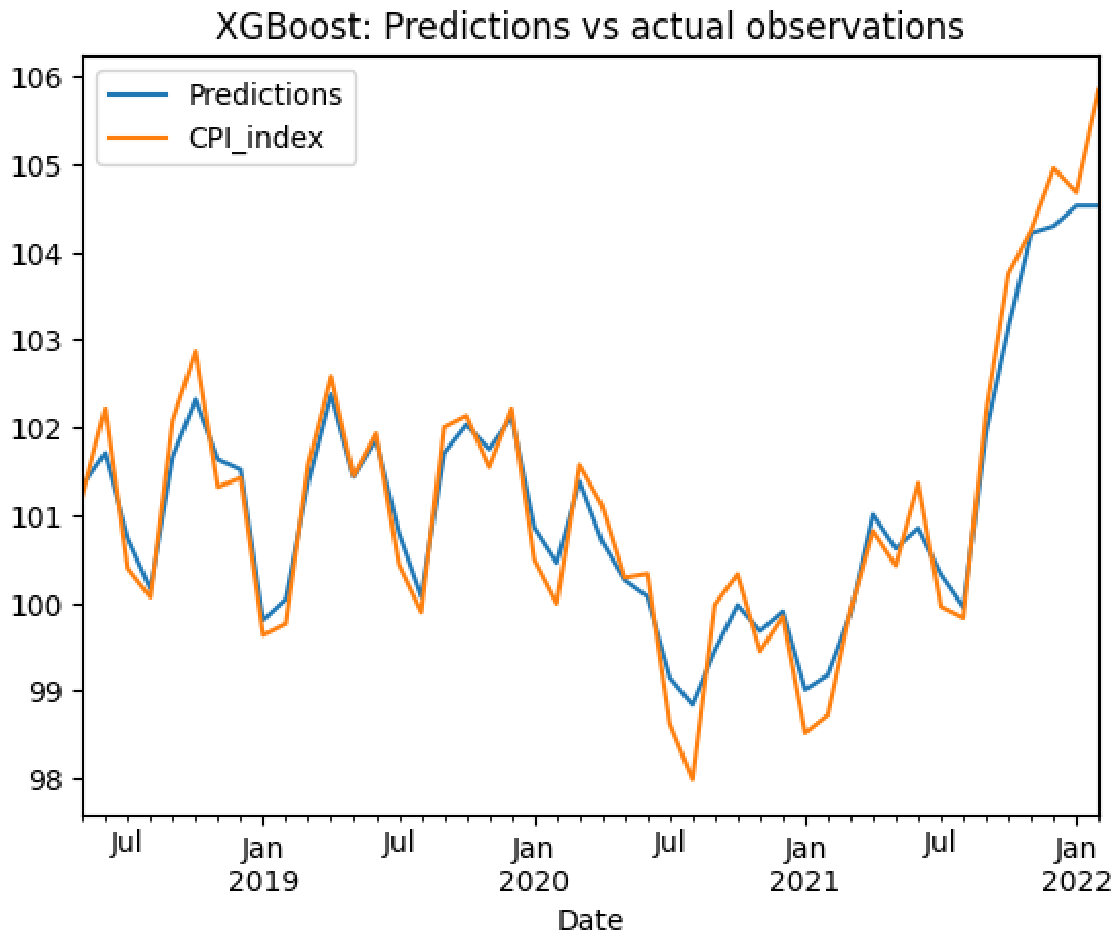
The first model that is built, utilized all relevant features discovered in the EDA, namely: CPI, import price, M3, Crude oil average price, Natural Gas Europe price and Capacity utilization. The model was computationally efficient, as its training time was significantly less than that of all other models.

Model Results

In terms of RMSE, the model outperformed both the LSTM ANN and the ARIMA(4, 0, 4), however, it is still well behind in comparison to the Persistence baseline. More specifically, the final model's RMSE is 3.419, significantly lower than the LSTM ANN, slightly lower than the ARIMA(4,0,4) and quite higher than that of the Baseline. If the Predictions vs Actual values line chart is examined, one will realize that until around 2022, the model performs better than all of the previous models, including the baseline, both in terms of RMSE (0.39 up to January 2022) and in capturing direction of changes. However, from 2022 onwards, the model is unable to perform at all. This may occur due to the exceptionally high and steep CPI increase, of unprecedented magnitude. Of course, it should not be omitted that the LSTM model was trained using cross validation, as opposed to the LSTM ANN model, and hence more data were utilized during training.

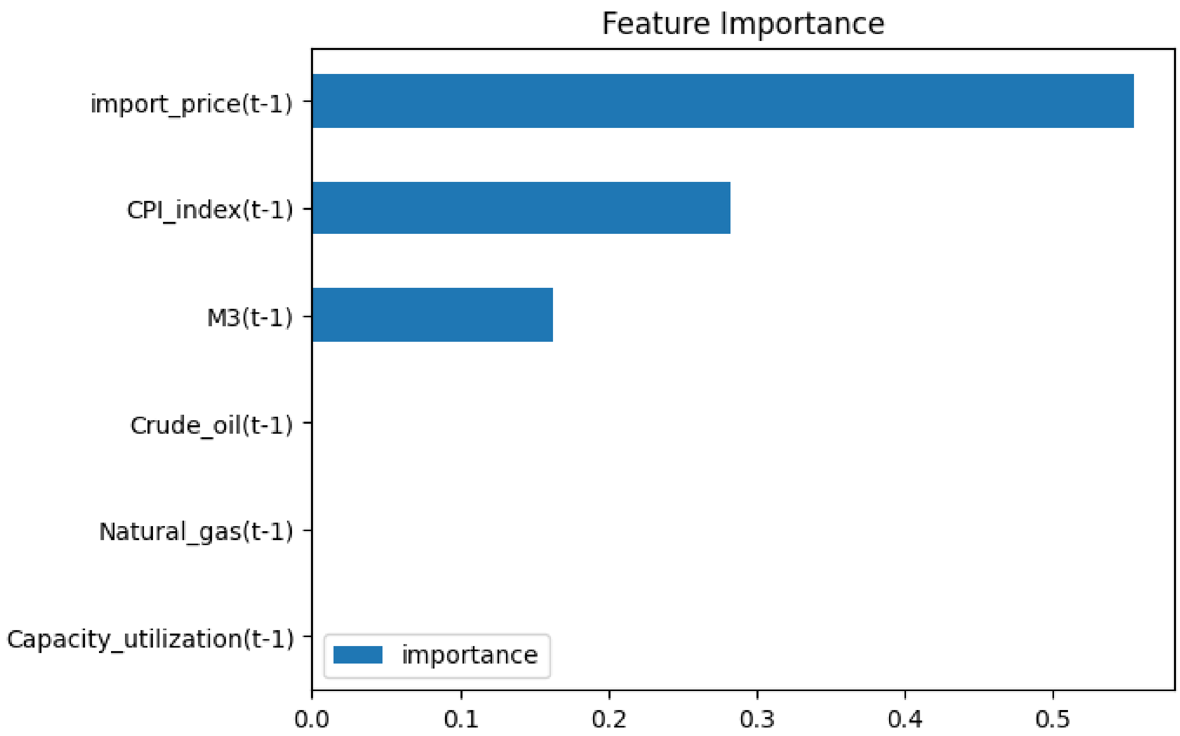


Line chart 24 – XGBoost CPI forecasts vs actual test set observations for model 1. Designed on python.



Line chart 25 – XGBoost CPI forecasts vs actual test set observations for model 1 (up to January 2022).

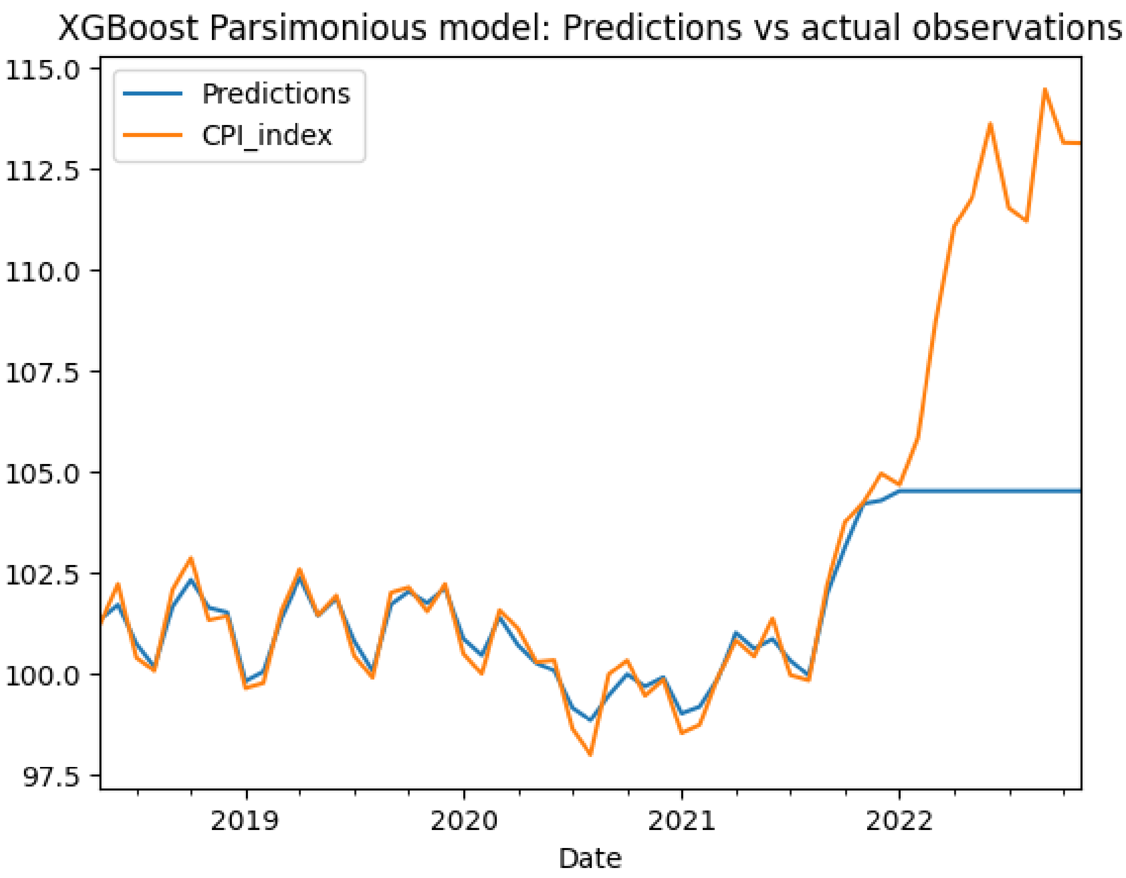
It is important to examine the variable importance for each of the input features that was forwarded to the algorithm. Out of all variables, import price, CPI and M3 seem to be significant with an importance rating of 0.55, 0.28 and 0.16 respectively. Natural Gas, Crude oil and Capacity utilization, presented an importance rating of only 0.00011, 0.001 and 0.000072 respectively (see bar chart 3).



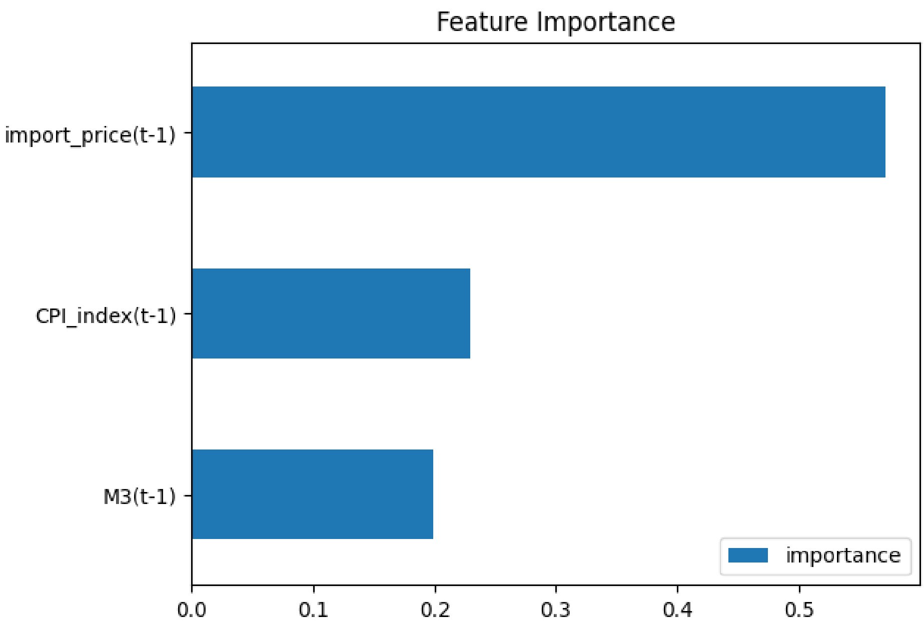
Bar chart 3 – XGBoost input features' importance. Designed on python.

Because Crude oil, Natural gas and Capacity utilization presented extremely low feature importance, it is deemed useful that a second, more parsimonious model is generated that excludes those features from the input list, as they may add useless complexity to the model.

The second model which only includes CPI, M3 and import price performs better and scores an RMSE of approximately 3.301, while maintaining high quality in terms of capturing the direction of changes of the CPI values (see line chart 26). All remaining features of the model are significant (see bar chart 4).



Line chart 26 – XGBoost best model's CPI forecasts vs actual test set observations. Designed on python.



Bar chart 4 – XGBoost best model's input features' importance. Designed on python.

4. Discussion

4.1 Summary of Findings

The primary objective of this research project has been to compare distinct forecasting methodologies for predicting inflation. Specifically, the study has examined four different approaches, namely, a persistence baseline model, an ARIMA model, a Long-Short Term Memory Artificial Neural Network (LSTM ANN), and an XGBoost model. The underlying proposition is that an ensemble method, represented by an XGBoost model, would present superior performance in forecasting short-term CPI values, due to its ability to combine simple base learners in generating predictions. In the context of macroeconomic forecasting, it is hypothesized that the simplicity inherent in a model like ARIMA, cannot capture the complexity that shapes inflation dynamics, as it solely relies on lagged values of the CPI. Conversely, a deep learning model such as the LSTM ANN, tends to be excessively complex and hence, this may result to overfitting, particularly when the available data is only limited. Thus, it is expected that XGBoost, that is basically an ensemble algorithm of simple decision trees (base learners) would exhibit the highest performance of all models.

The empirical exercise that is designed to compare the performance of those four models in predicting the CPI values, one month into the future, is performed on the same set of unseen data in order to promote the generalizability of the proposed models, but also to ensure the comparability of the models' performance.

The persistence model is a rather challenging benchmark to surpass, that exhibited the most favorable performance in terms of Root Mean Square Error (RMSE) with a value of 1.314. However, its inherent limitation lies in its failure to capture the directional changes in the CPI values, as it merely represents, a forward shift of the CPI trend line. The reason why the persistence model showcased such a low RMSE score, can be attributed to the fact that inflation is inherently a highly persistent measure, especially within short-term horizons. Nevertheless, when predicting inflation, it is detrimental that the direction of changes is forecasted accurately. Hence, although a persistence model is a challenging baseline, while at the same time effective, quick to built, and interpretable, it should not be considered a reliable method for actual inflation forecasting.

The ARIMA(4,0,4) also demonstrated favorable performance with an RMSE score of approximately 3.44. While it falls short of surpassing the baseline model and at the same time its adequacy is disputable due to the presence of temporal dependencies in the residuals, as illustrated in bar chart 1 and bar chart 2, ARIMA seems to outperform the LSTM model and competes with the XGBoost model. It is intriguing that the ARIMA model outperformed the LSTM model, even though it is essentially an agnostic approach, relying solely on lagged values of the target variable without incorporating additional inputs. The success of ARIMA relative to more complex ML models, may be attributed to the limited available data and hence algorithms such as XGBoost and LSTM, that are fundamentally data hungry, may struggle to sufficiently model the input and target variable relationships. In any case, ARIMA, although a traditional forecasting approach, can be very efficient in time-series forecasting, especially in the field of inflation forecasting, since its simplicity and moderate demand for data, can potentially model such relationships more accurately. Finally, it is important to highlight, that the ARIMA model, seems to capture the direction of changes much better than the baseline model, although the LSTM and XGBoost seem to perform better in this part.

In summary, ARIMA may serve as a viable forecasting method when confronted with only limited data or in the absence of additional input features beyond lagged values of the target variable, however its inability to capture turning points and precise directional changes is eminent.

The Long-Short-Term-Memory Artificial Neural Network model presented the lowest performance of all models as indicated by its RMSE value of 3.74. However, due to the limited amount of data, the model still demonstrates satisfactory results under the prevailing circumstances. Notably, the LSTM model exhibited the highest degree of overfitting in comparison to the other models and it probably did so due to the limited amount of available data. Despite this limitation, the LSTM model seems to predict the directional changes of the target variable rather accurately relative the baseline model and the other models. On top of that, the LSTM model generated the most accurate predictions for the period after January 2022 wherein the predicted values closely approximate the actual observations, despite the unprecedented spike in the CPI values during that time. However, it should be highlighted that in general, the LSTM model tends to underestimate the future CPI values in its forecasts, and this is perhaps the model’s most pronounced weakness. The reason for this constant underestimation of predictions can be attributed to the limited available data and the positive skewness observed in the original CPI data distribution (see appendix 6, line chart 39).

In conclusion, while the LSTM model demonstrates potential for effective inflation forecasting in large, data-rich datasets, and seems promising in capturing complex relationships even in cases of unprecedented scenarios, its performance is hindered by the constraints imposed by limited data and the skewed nature of the CPI data distribution.

Finally, the XGBoost presented the best performance among the three models investigated, aligning with the initial hypothesis of this research project. The model achieved an RMSE score of approximately 3.301, while it captured most accurately the direction of changes. On top of that, XGBoost exhibits the least amount of overfitting compared to the rest of the models. However, the model seems to deteriorate significantly after January 2022 when there is a CPI spike of unprecedented magnitude. It is interesting that up until that date, the model showcases the lowest RMSE score of all the four models (0.39) and the prediction trend line suggests optimal directional accuracy (see line chart 26). However, the XGBoost model seems completely incapable of forecasting the extreme spike of CPI values occurring after January 2022, and the predictions trend line seems to flatten out. There can be several reasons why an XGBoost model may cease to predict inflation accurately after a specific point in time when an unprecedented inflation spike occurs. One reason could be that the training data used to train the model does not include examples of such extreme inflation spikes and hence the model has not learned how to handle such scenarios. Another explanation could be that extreme inflation spikes may introduce new patterns that were not present in historical data used for training. Finally, it is eminent that like all other machine learning models, XGBoost has certain limitations, and it may not be able to capture all non-linear relationships and complex interactions between variables, or sudden shifts in data dynamics.

In summary, XGBoost proves to be a powerful approach for inflation forecasting, demonstrating excellent performance and generalizability on even moderately sized datasets. However, the model’s effectiveness in predicting extreme inflation spikes is constrained by the model’s weakness to handle such unprecedented events.

4.2 Limitations and future recommendations

Overall, the biggest limitation of this research project has been the limited size of available data. While there exist numerous macroeconomic variables that are potentially important predictors of inflation, the limited data size hampers the construction of more complex models and restricts the overall performance of the models developed. Complex machine learning algorithms such as Neural Networks or XGBoost require thousands or even hundreds of thousands of data points to achieve optimal performance. However, the available time series data for this project is only limited to a few hundred observations. This scarcity of data is a common problem in macroeconomic forecasting, as measurements in this domain are usually recorded on a monthly frequency and systematic data collection has begun relatively recently in most European countries, particularly after World War II.

As the EDA illustrated, features such as GDP, bond prices and yields and other commodity prices such as Soya beans prices, may add important information to the models, and therefore it is suggested that relevant future research includes such variables as inputs. Additionally, the inclusion of other variables that have not been explored in this research paper, such as exchange rates and other relevant microeconomic variables that are daily measured may lead to the generation of more accurate models. However, this also prerequires that the input features include a large and extensive series of past measurements.

Despite the limitation posed by the data size, this research project successfully enables comparisons among different models and offers insights into their predictive abilities. However, it is important to acknowledge that the generalizability of the findings may be limited compared to studies conducted on more extensive datasets. Future research endeavors should strive to access larger datasets in order to improve the accuracy and reliability of inflation forecasting models.

Moving on, the model performance was assessed using RMSE, one of the most widely used measures of performance. However, as previously illustrated, the direction of changes of the CPI values is also very important in assessing the predictive ability of a model and to a greater extent, in comparing different models' performance. Although in this research project, the direction of changes was referred to, using a visual approach, measurements such as directional accuracy may be helpful in assessing the performance of different models, so that more firm conclusions can be made.

Additionally, in designing the LSTM ANN and the XGBoost model, the architecture and hyperparameter combinations that can be tested are numerous. Further experimenting with the input features, their lookback window, and the combinations of hyperparameters and algorithm architectures can significantly improve the performance of the models.

Moreover, although it has been hypothesized that ensemble methods such as XGBoost may perform better in inflation forecasting, compared to traditional approaches or even deep learning models, the list of forecasting methods is exhaustive, and hence, ensemble methods such stacking, voting or bagging of different models, ensemble algorithms such as AdaBoost or Random forest, but also other types of deep learning methods such as Transformer Neural networks or other Traditional approaches such as the Phillips curve or Lasso Regression should also be compared in making more concrete conclusions.

Finally, despite the models in this research project outperforming the persistence baseline in terms of capturing the direction of changes in CPI values, none of them were able to achieve a lower RMSE score than the persistence baseline. The limitation of failing to surpass the performance of the persistence baseline highlights the challenge of accurately forecasting extreme and unprecedented inflation spikes or declines. Future research should focus on developing approaches that can improve model performance even in such extreme scenarios. This can be achieved through techniques such as feature engineering, data augmentation, and data engineering. Additionally, incorporating daily collected variables instead of monthly data may also enhance the models' ability to capture and forecast extreme inflation events. By addressing these limitations, researchers can strive to create more effective forecasting models that outperform the persistence baseline even in the face of extreme inflation dynamics.

4.3 Conclusion

Summing up, the purpose of this report has been to compare the performance of 3 different models in the context of inflation forecasting. Although none of the models attained a lower RMSE score than the persistence baseline, XGBoost, a widely used decision trees ensemble algorithm, presented significantly better results than both the ARIMA model, a traditional approach for inflation forecasting and the LSTM ANN, a popular time-series deep learning forecasting method. It should be highlighted however, that the persistence model is not considered a valid approach to forecast inflation, as it does not account for any directional changes at all. The baseline model is essentially a representation of a forward shift in the CPI values and is only selected in this research project as a baseline due to the inherited persistence in inflation dynamics, thus making a persistence model, a challenging benchmark.

It is important to mention that the ARIMA model, although it does not take into consideration inputs other than past CPI values, it seems to perform better than the LSTM ANN. This indicates that when dealing with a limited amount of data, simpler models may be more useful in macroeconomic forecasting in comparison to more advanced deep learning models that require a larger amount of data.

On the other hand, although neither the ARIMA nor the XGBoost model were able to perform optimally after the 2022 extreme inflation spike, the LSTM model, was able to forecast the inflation spike more effectively than its 'competitor' models and this is partly due to the greater ability of Neural Networks to model complex phenomena and relationships. Hence, it is suggested that training an LSTM model with large datasets, may lead to the ability of the model to capture unexpected inflation dynamics that other algorithms and approaches are not able to capture, even in unprecedented scenarios.

The XGBoost seemed to be the best performing model in this research project, as it showcased the smaller RMSE score of all three models. Additionally, XGBoost also seems to predict the direction of changes of the CPI values more accurately than all other models. However, a pronounced weakness of XGBoost appeared to be that it cannot forecast CPI values during the 2022 extreme inflation spike, and instead, the model seems to exhibit a flattening trend. Therefore, XGBoost might not need as much data as the LSTM model to effectively forecast inflation, as long as future inflation values are not too extreme in comparison to the training data values or such extreme scenarios are not completely unseen during training.

Overall, it is suggested that future research should focus on addressing several key aspects to enhance model performance. Firstly, it is imperative to work with significantly larger datasets in order to overcome the overfitting problem that was observed in the models of this research project. By incorporating more data points, models can better capture the underlying patterns and dynamics of inflation, leading to improved forecasting accuracy.

Additionally, having a larger dataset may allow for the inclusion of more variables, as illustrated in this report's EDA. Exploring more variables or the addition of daily recorded variables during modelling may also improve the performance of future models.

Moreover, future studies should compare and evaluate various ensemble methods, deep learning methods and traditional approaches when forecasting inflation. This comparative analysis may provide further insight into the strengths and weaknesses of different models, and hence, enable researchers and professional analyst to select the most suitable approach for their specific forecasting objectives.

Lastly, it might be valuable that future research considers additional performance metrics beyond RMSE, such as directional accuracy when comparing and evaluating different models. Assessing the models' ability to accurately predict the directional changes in CPI values is detrimental for practical forecasting applications. Hence, by incorporating a metric such as directional accuracy, researchers can gain a more complete understanding of the models' effectiveness.

By addressing these recommendations, future research in inflation forecasting can advance the field, leading to more accurate and reliable models to support decision-making in various industries and domains.

Appendices

Appendix 1 – Forecasting performance comparison for sixteen different models in twelve different time horizons.

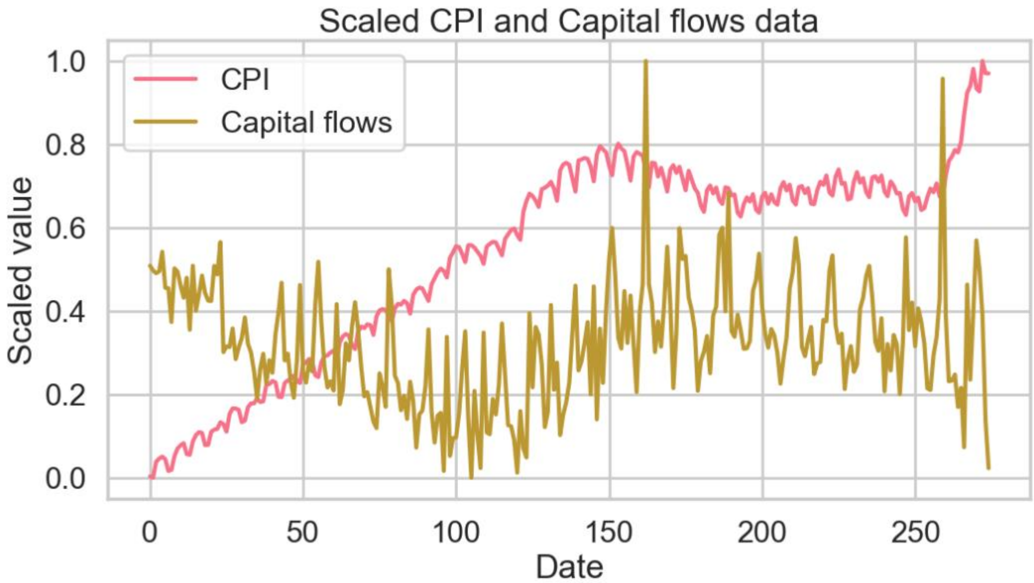
	forecast horizon (in months)											
Model	1	2	3	4	5	6	7	8	9	10	11	12
1 - Random walk	0.079**	0.123***	0.130***	0.160***	0.198***	0.196***	0.183***	0.169***	0.151***	0.140**	0.131**	0.135**
	(0.034)	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.003)	(0.010)	(0.020)	(0.038)	(0.028)
2 - Random walk (Atkeson and Ohanian)	0.103***	0.104***	0.105***	0.105***	0.107***	0.108***	0.109***	0.111***	0.113**	0.115**	0.117**	0.119***
	0.000	(0.001)	(0.002)	(0.003)	(0.004)	(0.003)	(0.006)	(0.009)	(0.011)	(0.012)	(0.012)	(0.008)
3 - ARMA	0.064	0.088***	0.090*	0.097**	0.103**	0.102**	0.102**	0.103*	0.104*	0.105*	0.107*	0.109**
	(0.120)	(0.008)	(0.060)	(0.040)	(0.023)	(0.030)	(0.045)	(0.051)	(0.060)	(0.057)	(0.056)	(0.049)
4 - VAR	0.062*	0.088***	0.090*	0.096**	0.105**	0.103**	0.101*	0.101*	0.103*	0.104*	0.107*	0.110*
	(0.091)	(0.008)	(0.051)	(0.043)	(0.014)	(0.030)	(0.054)	(0.064)	(0.067)	(0.061)	(0.058)	(0.054)
5 - Phillips curve (backward)	0.064*	0.092***	0.099**	0.101**	0.124***	0.127***	0.129***	0.120**	0.105**	0.102***	0.099*	0.096
	(0.070)	(0.004)	(0.031)	(0.019)	(0.006)	(0.008)	(0.008)	(0.015)	(0.032)	(0.006)	(0.068)	(0.174)
6 - Phillips curve (hybrid)	0.062	0.085**	0.101**	0.113**	0.141**	0.150**	0.153***	0.136**	0.118***	0.110***	0.102*	0.094
	(0.123)	(0.023)	(0.029)	(0.014)	(0.014)	(0.015)	(0.008)	(0.011)	(0.008)	(0.008)	(0.067)	(0.252)
7 - Factor model (direct forecast)	0.079***	0.081**	0.081**	0.093*	0.099*	0.100*	0.096*	0.099**	0.108**	0.113**	0.123**	0.135***
	(0.007)	(0.045)	(0.040)	(0.071)	(0.100)	(0.087)	(0.075)	(0.034)	(0.026)	(0.036)	(0.040)	(0.005)
8 - Factor model (iterated forecast)	0.052	0.074	0.080*	0.084**	0.101***	0.096*	0.097	0.100*	0.103*	0.106*	0.109**	0.112**
	(0.645)	(0.189)	(0.084)	(0.014)	(0.003)	(0.079)	(0.119)	(0.095)	(0.076)	(0.054)	(0.045)	(0.050)
9 - Factor model (direct forecast, targeted)	0.068*	0.074	0.073	0.086***	0.093*	0.090**	0.108**	0.081	0.099**	0.096	0.098	0.088
	(0.085)	(0.155)	(0.250)	(0.009)	(0.083)	(0.031)	(0.022)	(0.401)	(0.048)	(0.115)	(0.168)	(0.379)
10 - Factor model (iterated forecast, targeted)	0.07	0.092*	0.089**	0.100***	0.111***	0.102**	0.100**	0.098*	0.097	0.103*	0.106**	0.105*
	(0.127)	(0.074)	(0.040)	(0.007)	(0.002)	(0.018)	(0.047)	(0.086)	(0.125)	(0.060)	(0.039)	(0.094)
11 - Elastic net	0.080***	0.085***	0.085***	0.089***	0.090***	0.087**	0.085*	0.090**	0.092*	0.092*	0.099**	0.106***
	(0.001)	(0.001)	(0.004)	(0.005)	(0.003)	(0.017)	(0.052)	(0.045)	(0.094)	(0.073)	(0.012)	(0.008)
12 - LASSO	0.081***	0.085***	0.083***	0.087***	0.091***	0.088**	0.086**	0.091*	0.091	0.087	0.103***	0.108***
	(0.001)	(0.001)	(0.005)	(0.003)	(0.002)	(0.014)	(0.050)	(0.059)	(0.143)	(0.209)	(0.007)	(0.005)
13 - Ridge regression	0.076**	0.080**	0.082**	0.084**	0.086**	0.084*	0.084	0.084	0.087	0.09	0.093*	0.095
	(0.014)	(0.033)	(0.044)	(0.045)	(0.050)	(0.078)	(0.113)	(0.190)	(0.197)	(0.140)	(0.092)	(0.106)
14 - Random forest	0.070**	0.078**	0.079*	0.082**	0.083*	0.081*	0.081	0.082	0.084	0.086	0.088	0.09
	(0.014)	(0.023)	(0.055)	(0.049)	(0.064)	(0.098)	(0.161)	(0.282)	(0.346)	(0.253)	(0.231)	(0.238)
15 - Quantile regression forest	0.072**	0.079**	0.081*	0.083**	0.084*	0.082*	0.082	0.083	0.085	0.086	0.089	0.09
	(0.015)	(0.030)	(0.053)	(0.047)	(0.062)	(0.099)	(0.143)	(0.238)	(0.306)	(0.261)	(0.238)	(0.259)
16 - Focus survey	0.055	0.065	0.068	0.069	0.072	0.074	0.076	0.079	0.082	0.082	0.084	0.085
	-	-	-	-	-	-	-	-	-	-	-	-
Number of observations	95	94	93	92	91	90	89	88	87	86	85	84

Table 6 - ML models for inflation forecasting RMSE comparison (by Araujo, G. S. & Wagner, P. G., 2020).

Appendix 2 – Insignificant variables for inflation prediction

Capital flows.

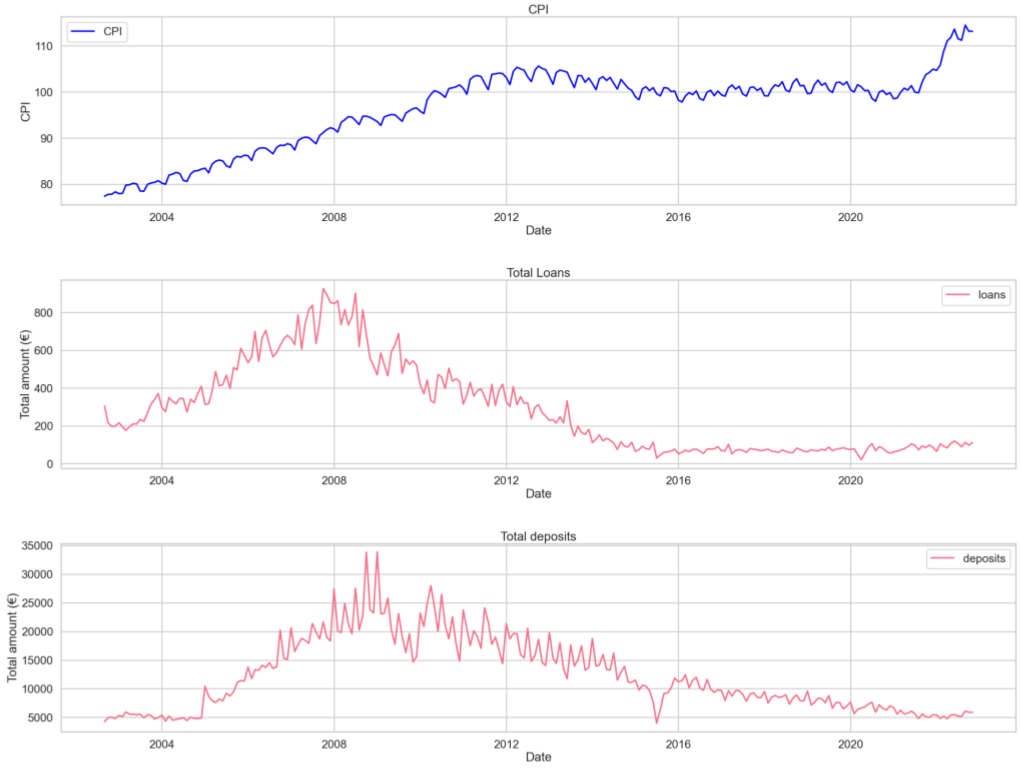
Capital flows and the CPI trend lines do not seem to be strongly related.



Line chart 27 - Scaled CPI values and Capital flows trend lines

Total deposits and total loans.

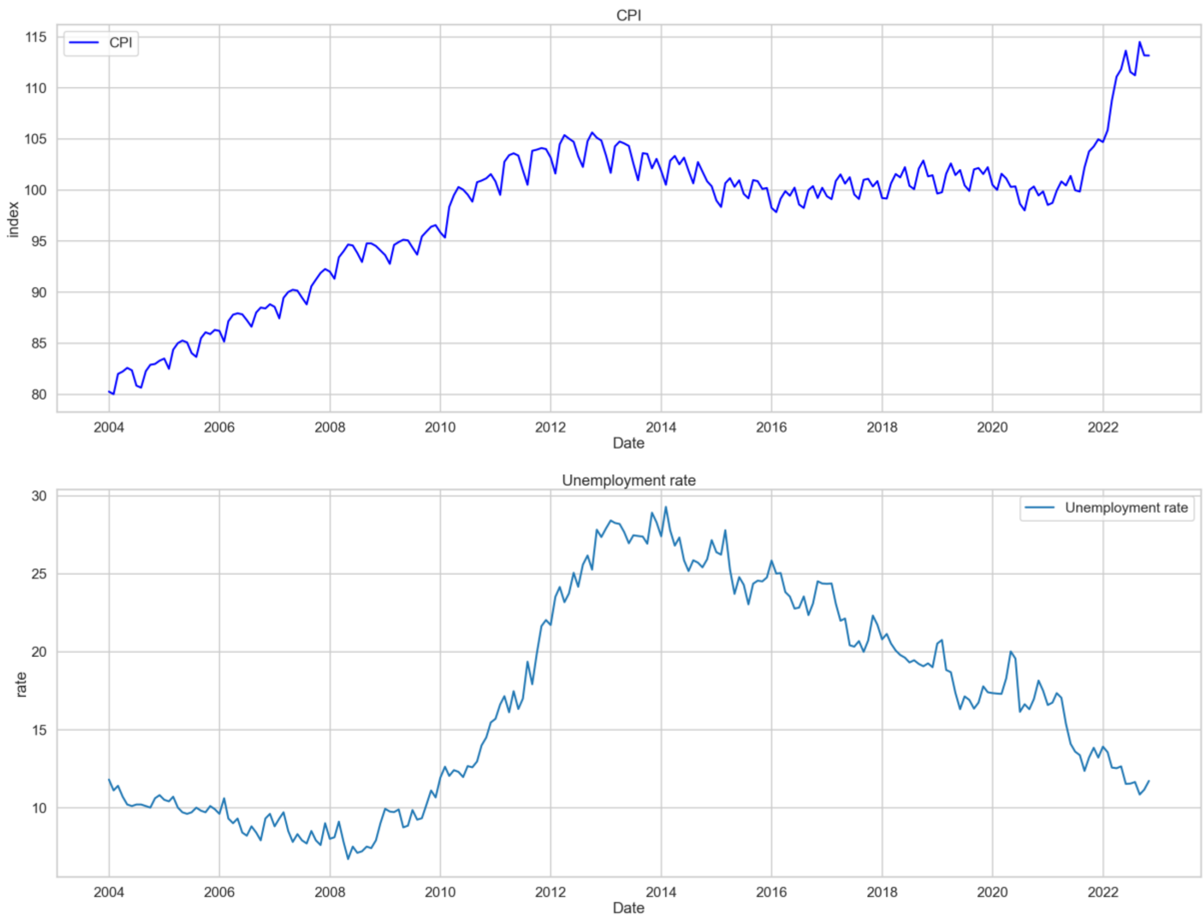
Total deposits and total loans and the CPI trend lines do not seem to be strongly related.



Line chart 28 - CPI values, total deposits, and total loans trend lines.

Unemployment rate

Unemployment rate and the CPI trend lines do not seem to be strongly related.



Line chart 29 - CPI values, and unemployment rate trend lines.

Appendix 3 – Significantly correlated energy variables

All couples of variables bellow have a Pearson's coefficient larger than 0.9 and hence it is deemed that they are significantly correlated. Variables with such a high correlation may lead to multicollinearity issues during modeling, and hence, highly correlated variables are filtered so that only the most relevant variables are used, according to theory and the exploratory data analysis.

Variable 1	Variable 2
Crude oil, average	Crude oil, Brent
Crude oil, average	Crude oil, Dubai
Crude oil, average	Iron ore, cfr spot
Crude oil, average	Copper
Crude oil, average	Platinum
Crude oil, Brent	Crude oil, average
Crude oil, Brent	Crude oil, Dubai
Crude oil, Brent	Iron ore, cfr spot
Crude oil, Brent	Copper
Crude oil, Brent	Platinum
Crude oil, Dubai	Crude oil, average
Crude oil, Dubai	Crude oil, Brent
Crude oil, Dubai	Iron ore, cfr spot

Crude oil, Dubai	Copper
Crude oil, Dubai	Platinum
Tea, avg 3 auctions	Tea, Colombo
Tea, avg 3 auctions	Tea, Mombasa
Tea, avg 3 auctions	Tobacco, US import u.v.
Tea, Colombo	Tea, avg 3 auctions
Tea, Colombo	Banana, US
Tea, Colombo	Tobacco, US import u.v.
Tea, Mombasa	Tea, avg 3 auctions
Groundnut oil	Soybeans
Groundnut oil	Soybean oil
Groundnut oil	Maize
Palm oil	Soybeans
Palm oil	Soybean oil
Soybeans	Groundnut oil
Soybeans	Palm oil
Soybeans	Soybean oil
Soybeans	Soybean meal

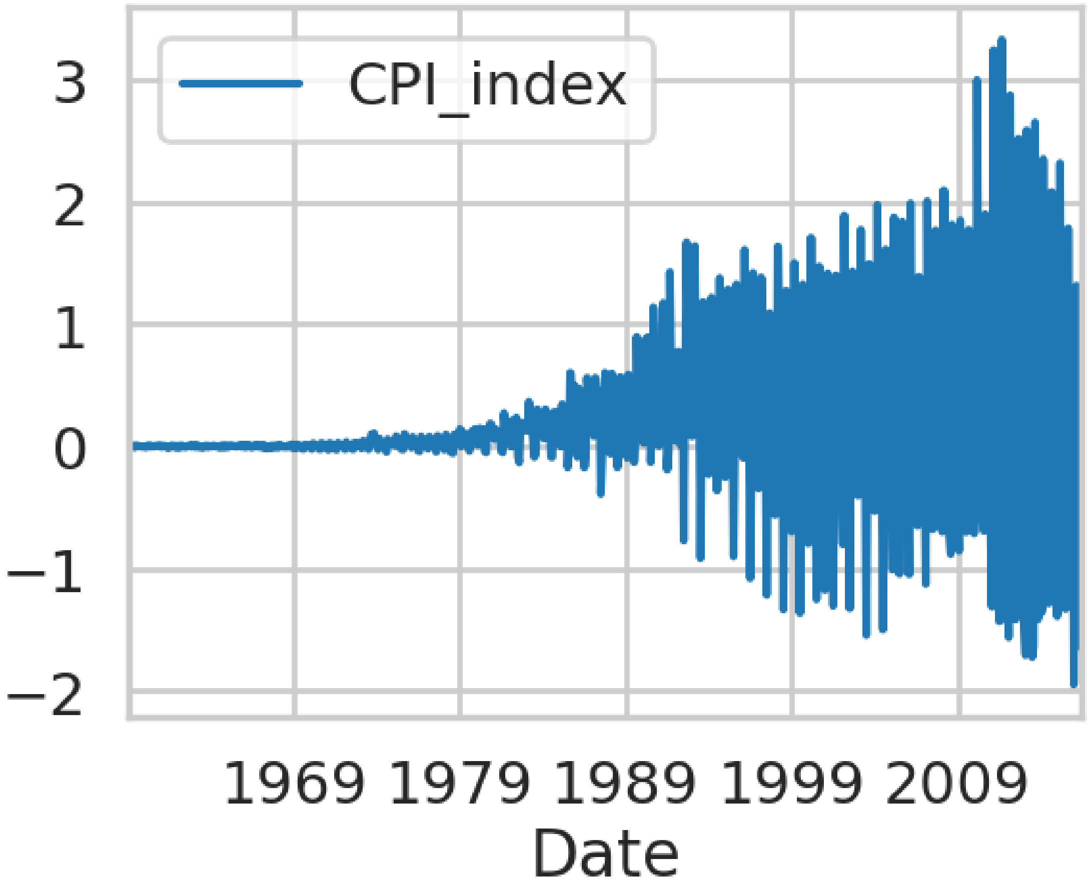
Soybeans	Maize
Soybeans	Wheat, US HRW
Soybean oil	Groundnut oil
Soybean oil	Palm oil
Soybean oil	Soybeans
Soybean oil	Maize
Soybean oil	Wheat, US HRW
Soybean oil	Urea
Soybean meal	Soybeans
Maize	Groundnut oil
Maize	Soybeans
Maize	Soybean oil
Maize	Wheat, US HRW
Wheat, US HRW	Soybeans
Wheat, US HRW	Soybean oil
Wheat, US HRW	Maize
Banana, US	Tea, Colombo
Banana, US	Beef

Banana, US	Chicken
Banana, US	Tobacco, US import u.v.
Banana, US	Gold
Beef	Banana, US
Beef	Gold
Chicken	Banana, US
Tobacco, US import u.v.	Tea, avg 3 auctions
Tobacco, US import u.v.	Tea, Colombo
Tobacco, US import u.v.	Banana, US
Logs, Malaysian	Sawnwood, Malaysian
Sawnwood, Malaysian	Logs, Malaysian
TSP	Urea

Table 7 - Highly correlated Commodity Price variables

Appendix 4 – Making CPI data stationary.

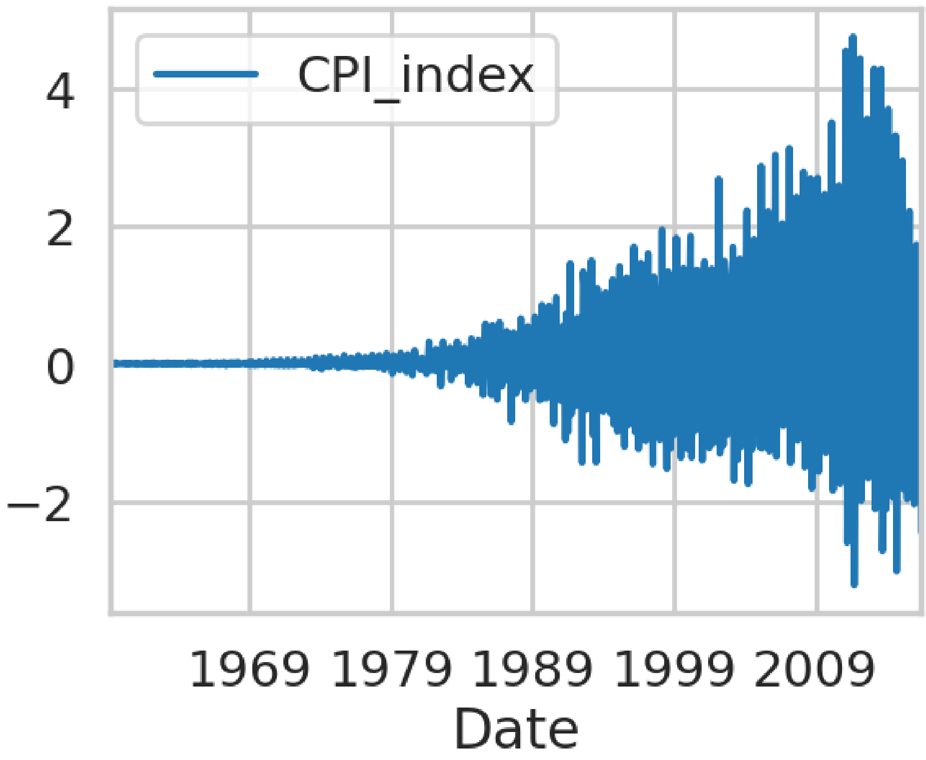
First difference



Line chart 30 – First difference of CPI values

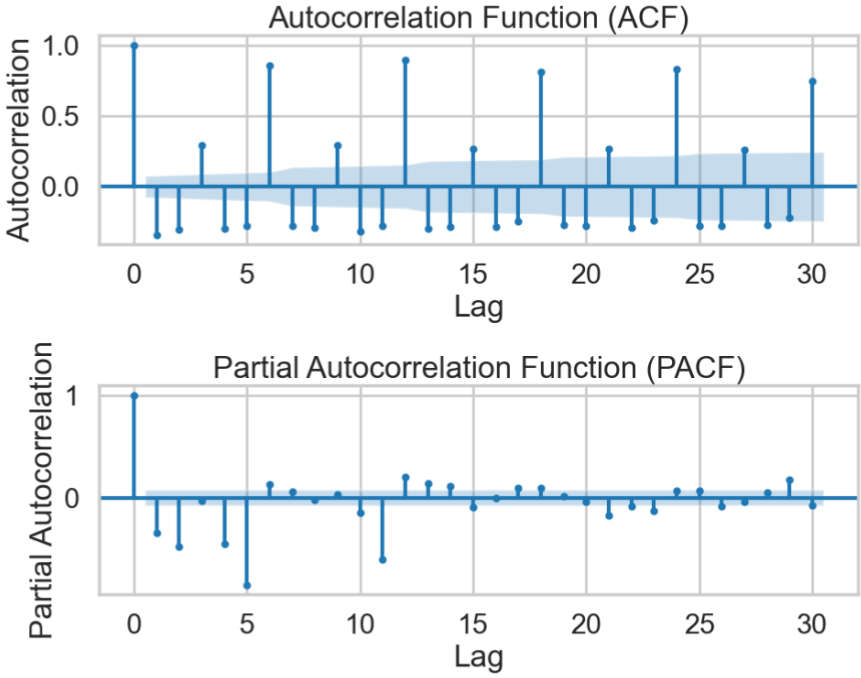
After the first differences of the CPI values were calculated, it was found that the data is still not stationary, as it is eminent in line chart 30, since as time progresses, the fluctuations of the transformed CPI values increase significantly in magnitude and as suggested by the KPSS test ($p < 0.05$) and ADF test ($p > 0.05$) for stationarity.

Second difference



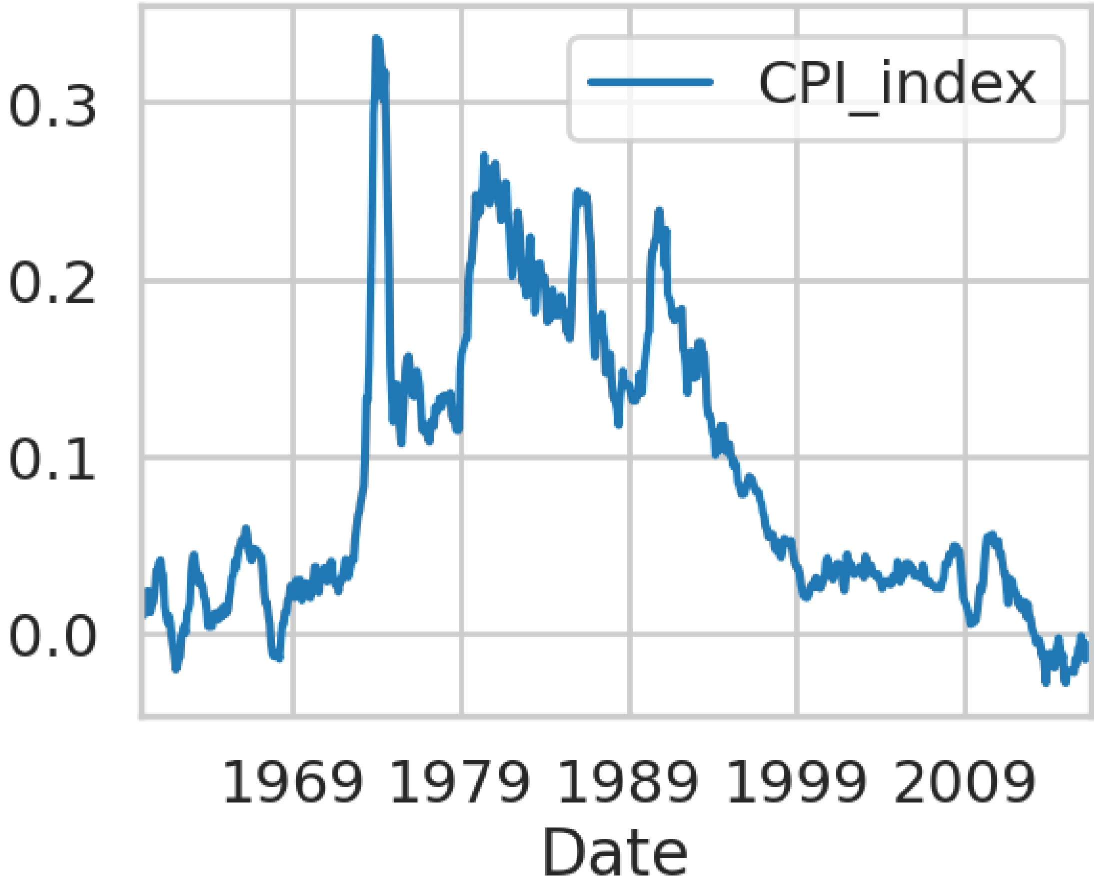
Line chart 31 – Second difference of CPI values

After the second differences for the CPI values were calculated, it was found that the data is still not stationary, as it is eminent in line chart 31, since as time progresses, the fluctuations of the CPI values increase in magnitude. However, it was suggested by the KPSS test ($p > 0.05$) and ADF ($p < 0.05$) test for stationarity that the time series is stationary. In resolving the disagreement between the stationarity checks, the partial autocorrelation and autocorrelation plots were generated (bar chart 5). It is eminent that the order of the ARIMA cannot be estimated using the second difference transformation of the CPI values, since most lags are significant.



Bar chart 5 – Autocorrelation and Partial autocorrelation plot of second differenced CPI values

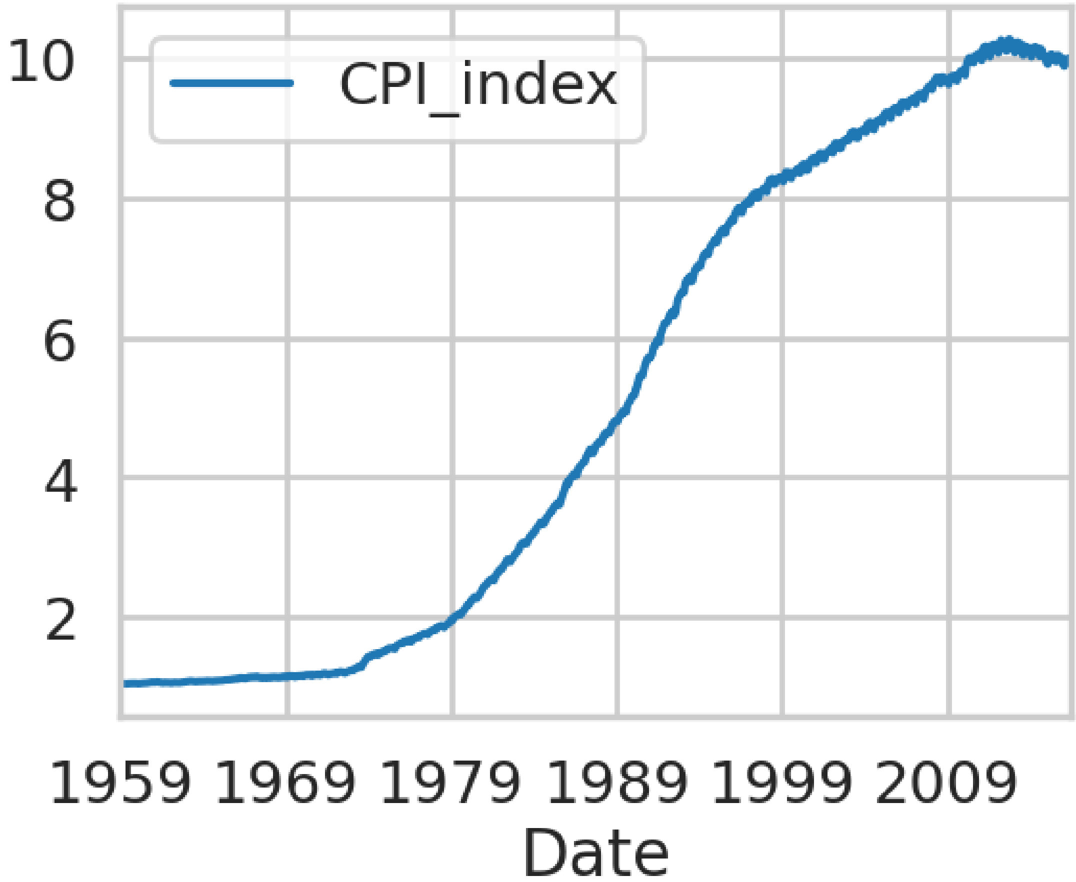
Seasonal percentage change (annually)



Line chart 33 –Annual Percentage change of CPI values trend line.

After the annual percentage change of the CPI values was calculated, it was found that the data is still not stationary, as it is eminent in line chart 33, that the transformed CPI trend line mean and variance is not stable and as suggested by the KPSS test ($p < 0.05$) and ADF test ($p > 0.05$) for stationarity.

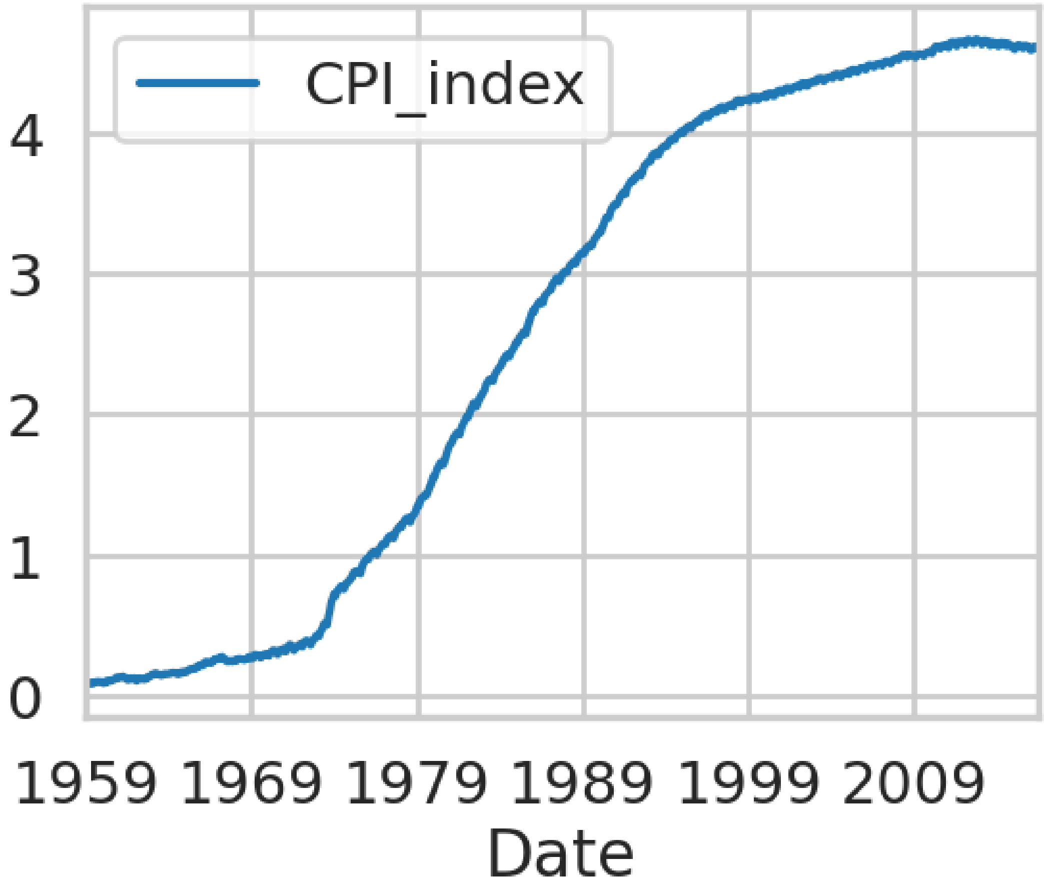
Logarithmic Transformation



Line chart 34 –Logarithmically transformed CPI values trend line.

The logarithmic transformation of the CPI values simply changed the scale of the values and did not affect the distribution of the observations. The time series is not stationary.

Square root Transformation



Line chart 35 –Square root of CPI values trend line.

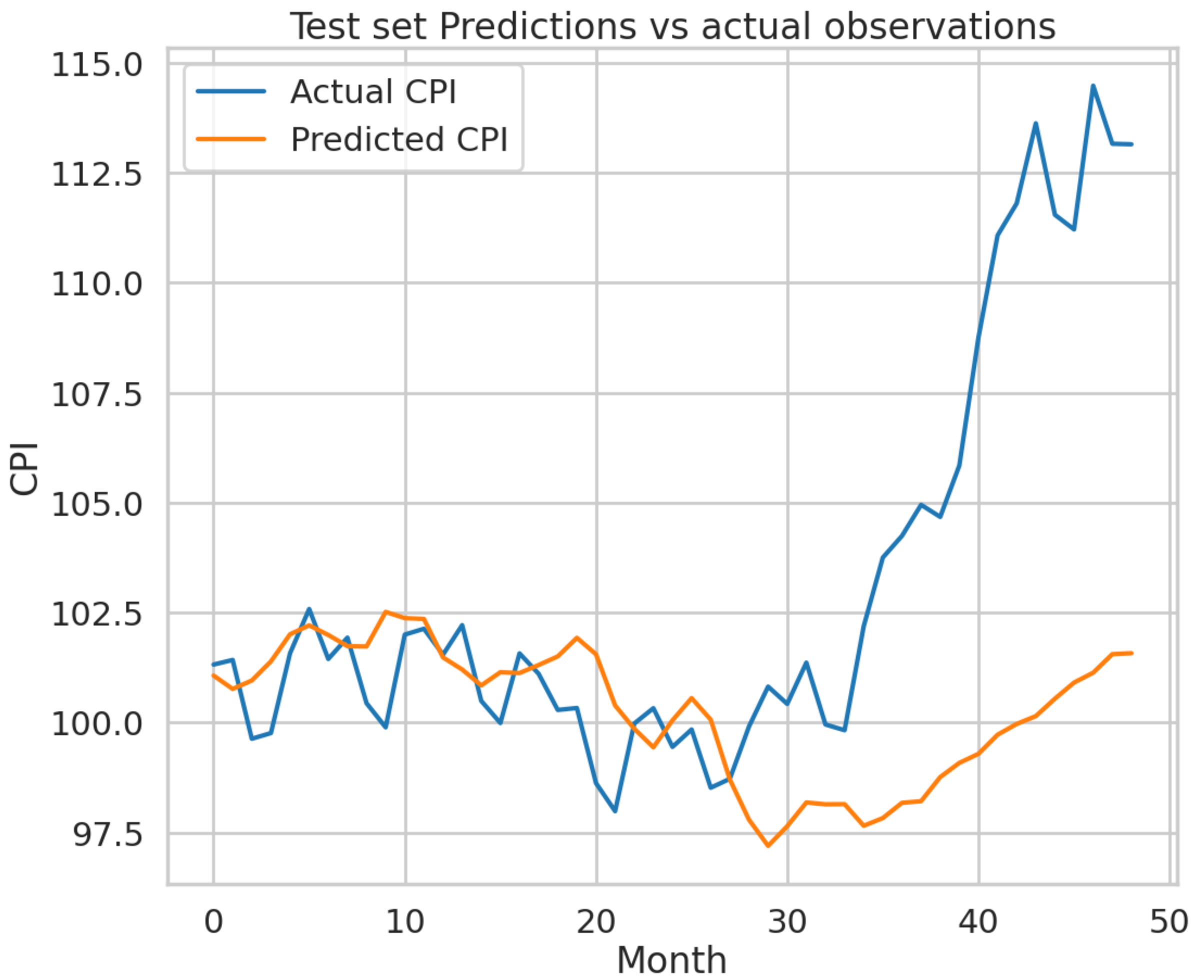
Square root transformation of the CPI values simply changed the scale of the values and did not affect the distribution of the observations. The time series is not stationary.

Appendix 5 – LSTM Architecture adjustments and Hyperparameter tuning.

The hyperparameters for all models bellow have been tuned (using Ray Tune) so that each model performs the best.

Model 1: With 1 fully connected layer (linear transformation) and no dropout layer

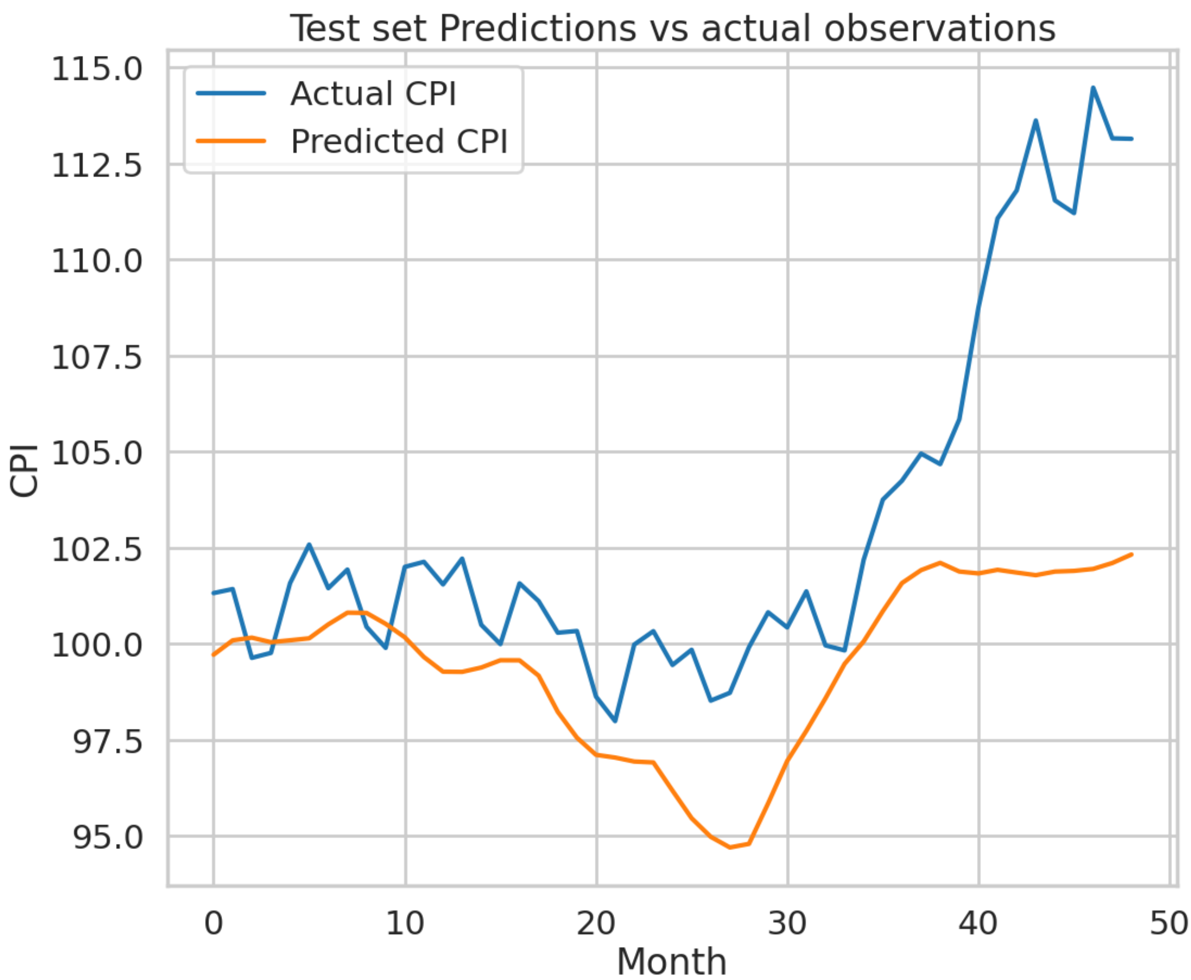
Best hyperparameters: Batch size = 64, lr = 0.001, Hidden layers = 40, number of stacked layers = 3, epochs = 1000.
RMSE = 5.55.



Line chart 36 –Model 1 predictions vs CPI actual values trend line. Designed on Python.

Model 2: With 2 fully connected layers and no dropout layer

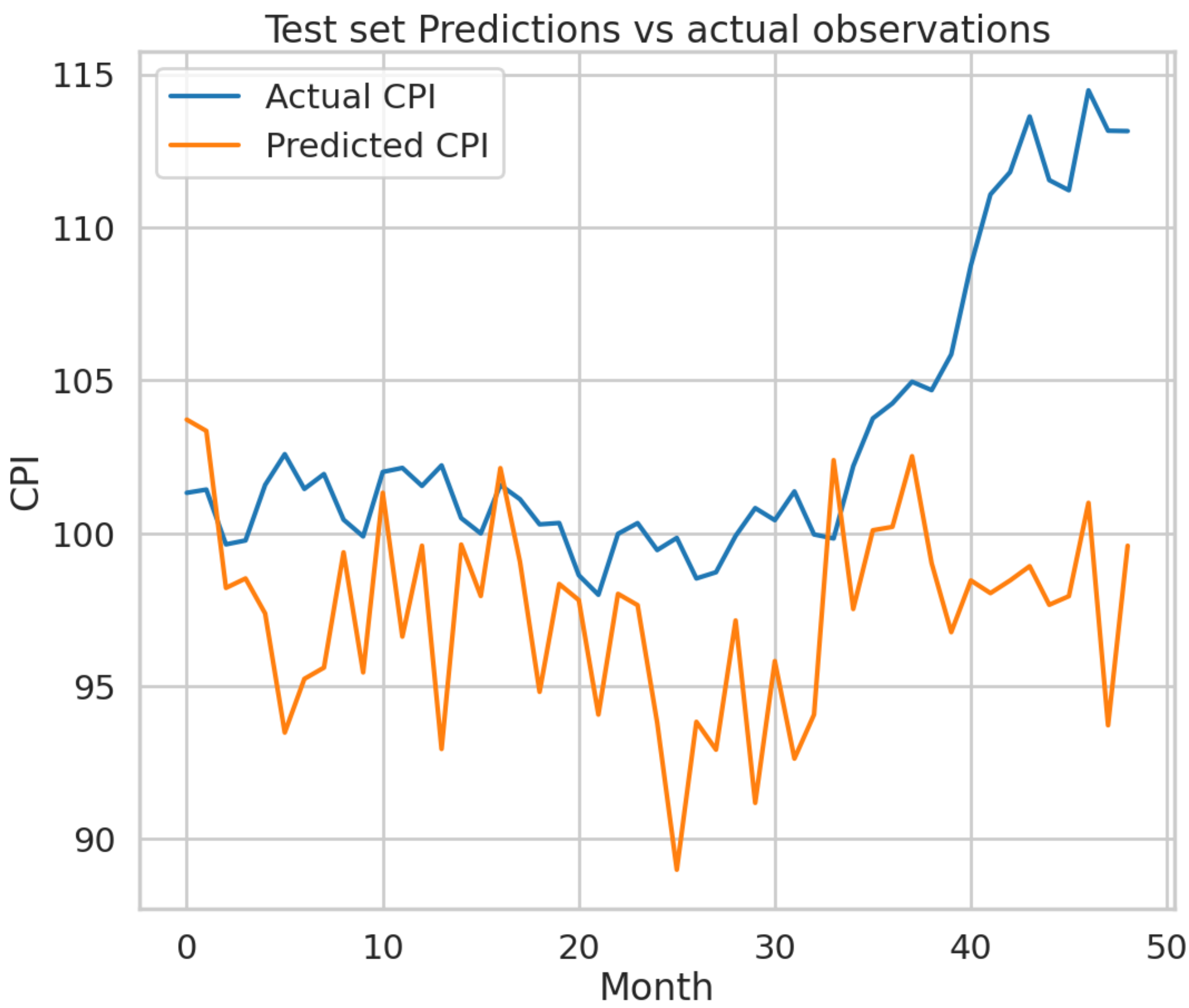
Best Hyperparameters: Batch size = 64, lr = 0.001, Hidden layers = 40, number of stacked layers = 4, epochs =1000.
RMSE = 7.27



Line chart 37 –Model 2 predictions vs CPI actual values trend line. Designed on Python.

Model 3: With 2 fully connected layers and a dropout layer

Best Hyperparameters: Batch size = 64, lr = 0.001, Hidden layers = 60, stacked layers = 3, epochs = 1000.
RMSE = 7.99

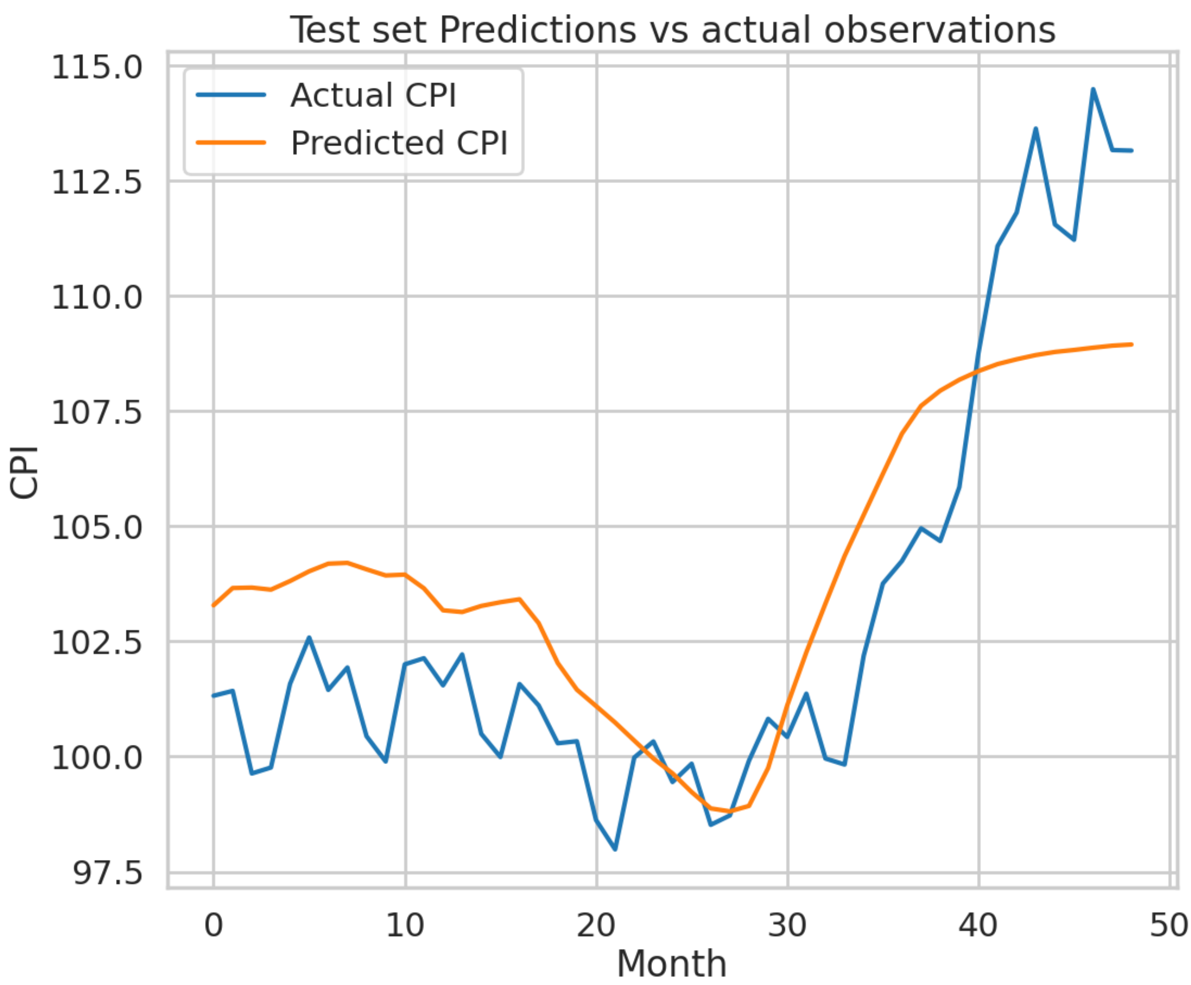


Line chart 37 –Model 3 predictions vs CPI actual values trend line. Designed on Python.

Model 4: With 3 fully connected layers and no dropout layer and no activation function

Best hyperparameters: Batch size = 64, lr = 0.001, Hidden layers = 60, stacked layers = 6, epochs = 1000.
RMSE = 3.81

This model, although it is simpler and more parsimonious than the best performing model, it does not seem to capture at all the changes of the CPI values direction.

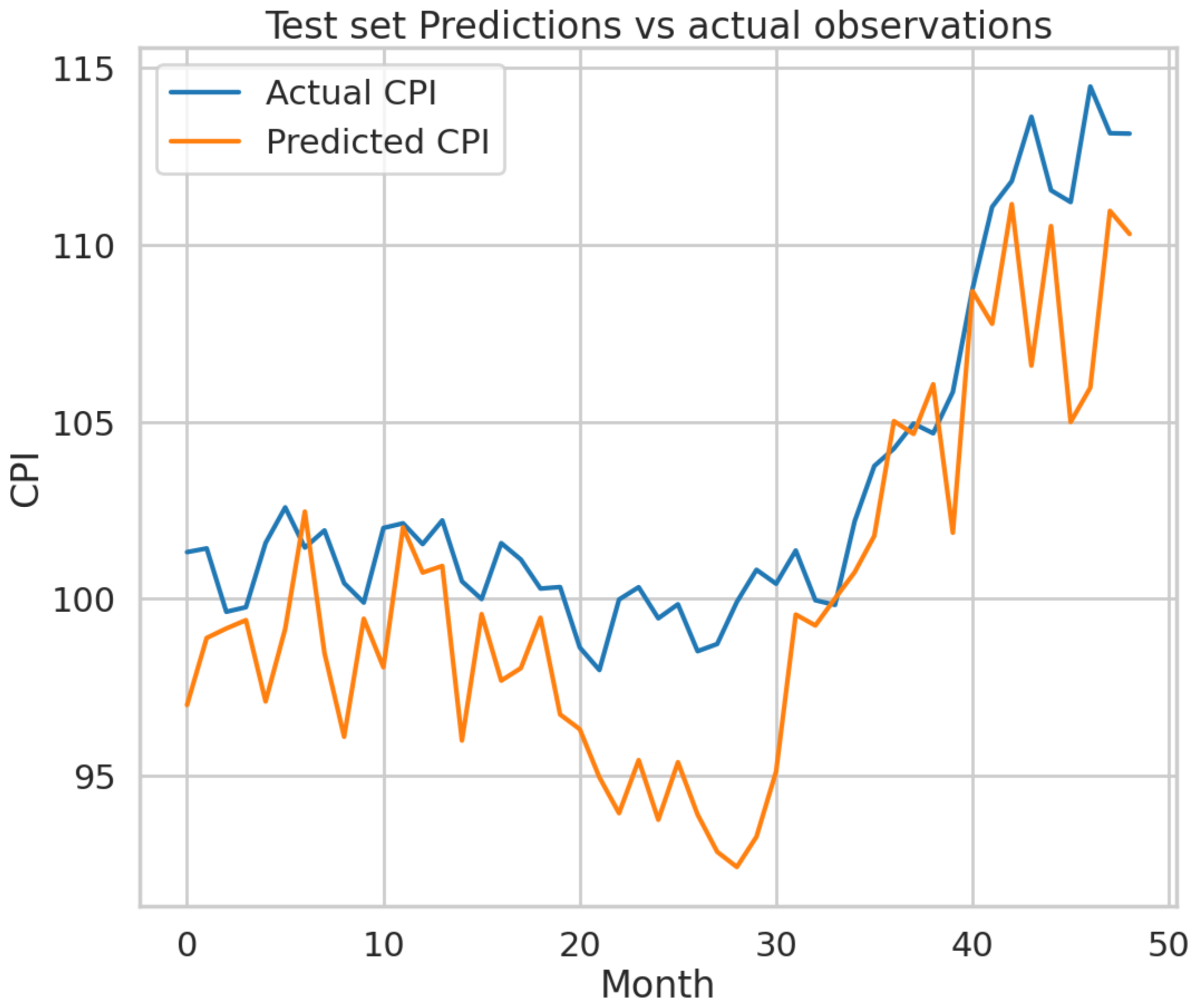


Line chart 37 –Model 4 predictions vs CPI actual values trend line. Designed on Python.

Model 5: With 3 fully connected layers, a dropout layer, and no activation function.

Best hyperparameters: Batch size = 64, lr = 0.001, Hidden layers = 60, stacked layers = 6, epochs = 1000.
RMSE = 3.79

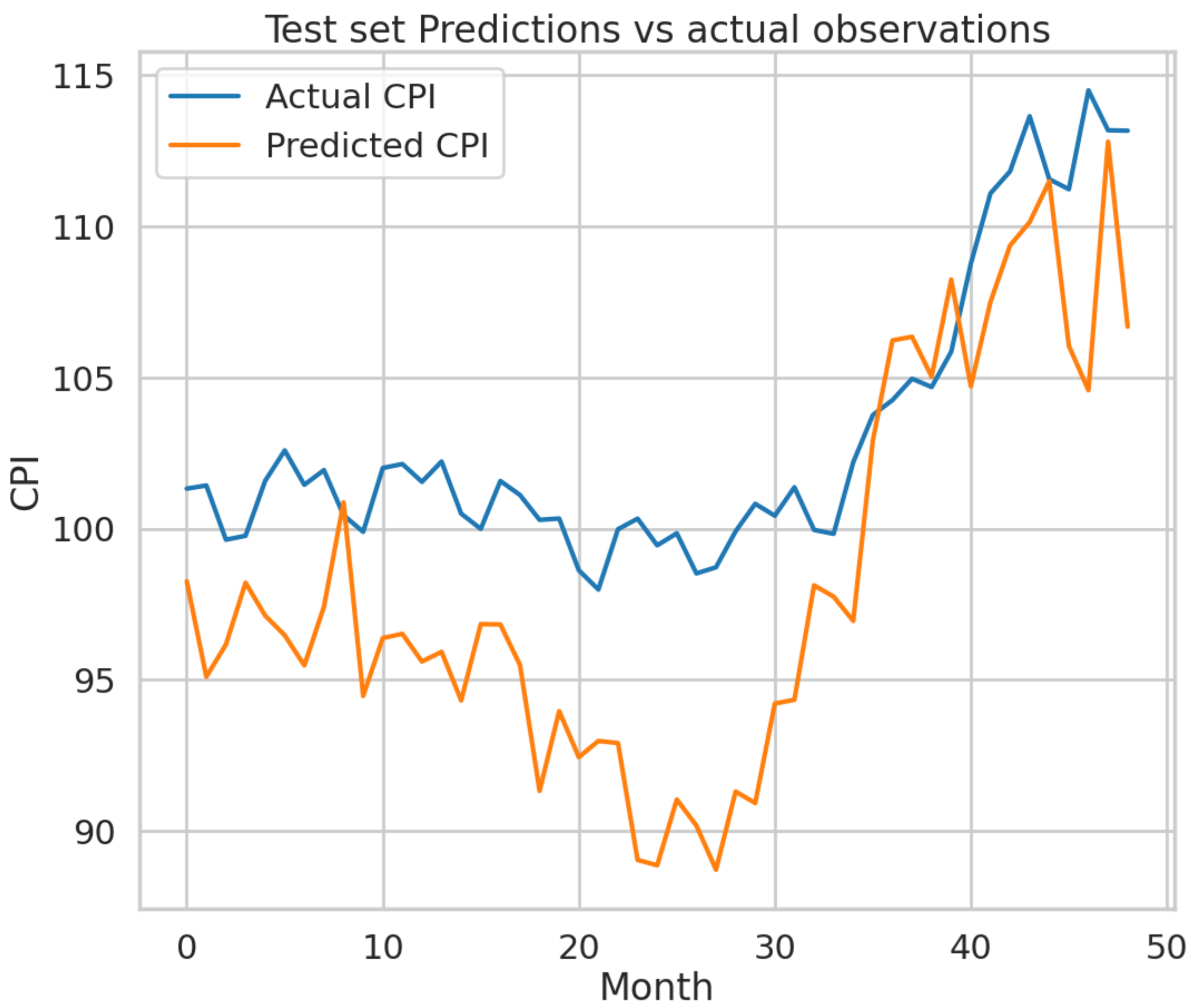
This model, does not include an activation function and hence is simpler than the suggested model, however its performance has slightly dropped.



Line chart 37 –Model 5 predictions vs CPI actual values trend line. Designed on Python.

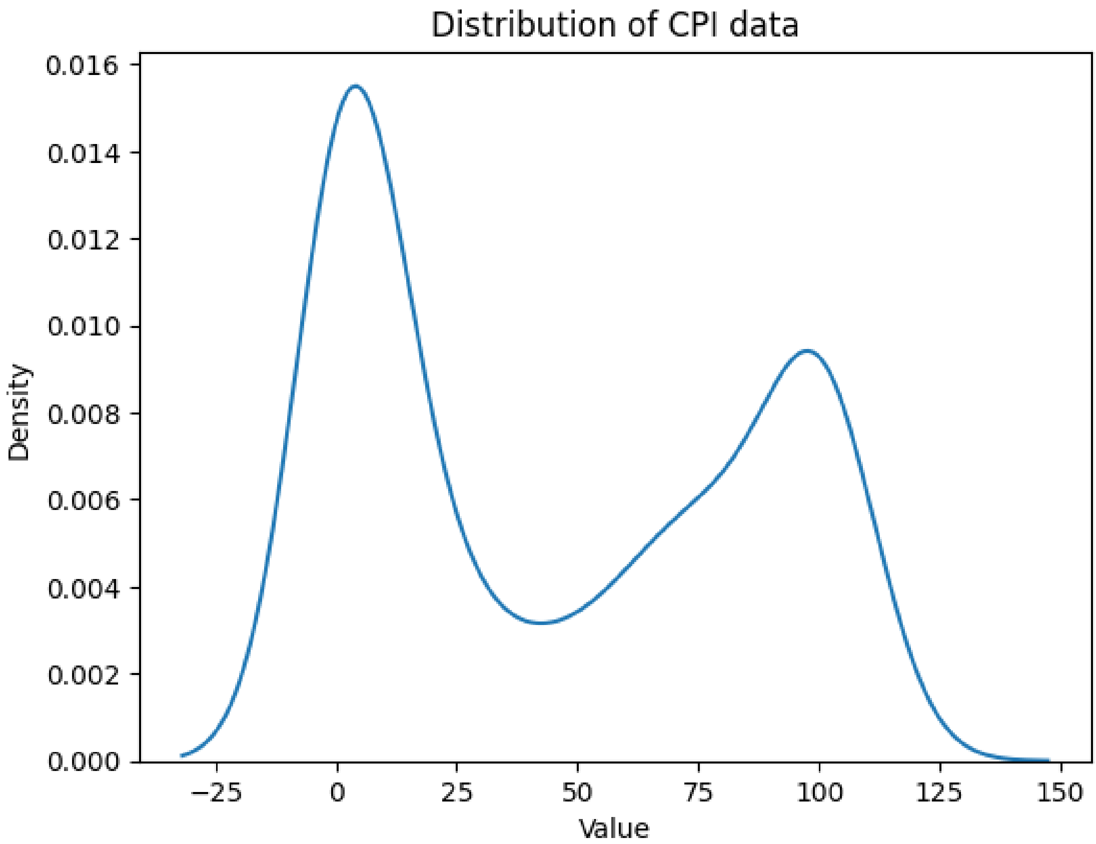
Model 6: With 4 fully connected layers, a dropout layer, and the ReLU activation function.

Best hyperparameters: Batch size = 64, lr = 0.001, Hidden layers = 60, stacked layers = 6, epochs = 1000.
RMSE = 5.86.



Line chart 38 –Model 6 predictions vs CPI actual values trend line. Designed on Python.

Appendix 6 – CPI data distribution



Line chart 39 –Distribution of CPI data. Designed on Python.

It is eminent in line chart 39(kernel density plot), that most CPI values are lower than 25, while there is also a spike of CPI with values around the 100 units. This suggests that the data can be considered positively skewed and hence, the results of the ML models might be influenced by such a distribution.

References

1. A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional LSTM and other neural network architectures,” Neural Netw., vol. 18, nos. 5–6, pp. 602–610, Jul./Aug. 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608005001206
2. A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, “A novel connectionist system for unconstrained handwriting recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 5, pp. 855–868, May 2009.
3. A. Graves. (Aug. 2013). “Generating sequences with recurrent neural networks.” [Online]. Available: http://arxiv.org/abs/1308.0850
4. A. J. Robinson and F. Fallside, The Utility Driven Dynamic Error Propagation Network. Cambridge, U.K.: Univ. Cambridge Department of Engineering, 1987.
5. Aggarwal, C. C. (2013). Outlier analysis (2nd ed.). Springer.
6. Alam, Kazi Md Rokibul, Nazmul Siddique, and Hojjat Adeli. "A dynamic ensemble learning algorithm for neural networks." *Neural Computing and Applications* 32 (2020): 8675-8690.
7. Ang, A., & Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics, 50(4),
8. Ang, Andrew, Geert Bekaert, and Min Wei. "Do macro variables, asset markets, or surveys forecast inflation better?." *Journal of monetary Economics* 54.4 (2007): 1163-1212.
9. Araujo, Gustavo Silva, and Wagner Piazza Gaglianone. "Machine learning methods for inflation forecasting in Brazil: new contenders versus classical models." *Latin American Journal of Central Banking* 4.2 (2023): 100087.
10. Barkan, Oren, et al. "Forecasting CPI Inflation Components with Hierarchical Recurrent Neural Networks." *International Journal of Forecasting* (2022).
11. Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. ACM Computing Surveys, 41(3), 1-52.
12. Belke, A., & Klose, J. (2014). Does the ECB rely on a Taylor rule? - comparing ex-post with real-time data. Empirical Economics, 46(1), 1-20.
13. Benediktsson J, Sveinsson J, Ersoy O and Swain P (1994). Parallel consensual neural networks with optimally weighted outputs. In: Proceedings of the World Congress on Neural Networks III. Lawrence Erlbaum Associates: Mahwah, NJ, USA, pp 129±137.

14. Benjamin Lindenmann, Timo Müller, Hannes Vietz, Nasser Jazdi, Michael Weyrich, A survey on long short-term memory networks for time series prediction, 2020.
15. Berdell, John F. 1996. "Innovation and Trade: David Hume and the Case for Freer Trade." Hist. Polit. Econ. 28 (Spring): 107–26.
16. Blanchard, O., 2000, "Macroeconomics (2nd ed.)", Prentice Hall.
17. Blaug, Mark. 1985. Economic Theory in Retrospect. 4th ed. Cambridge: Cambridge Univ. Press.
18. Capistrán, Carlos, and Manuel Ramos-Francia. "Does inflation targeting affect the dispersion of inflation expectations?." *Journal of Money, Credit and Banking* 42.1 (2010): 113-134.
19. Chen, T. and Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. arXiv preprint arXiv:1603.02754.
20. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable Tree boosting system.
21. Cheng, Kayan, Naijing Huang, and Zhentao Shi. "Survey-based forecasting: To average or not to average." *Behavioral Predictive Modeling in Economics* (2021): 87-104.
22. Choudhry, M. (2013). An introduction to bond markets. John Wiley & Sons.
23. Clemen R (1989). Combining forecasts: a review and annotated bibliography with discussion. Int J Forecast 5: 559±608.
24. Collins Dictionary, 1970
25. Comley, P., 2015, Inflation Matters: Inflationary wave theory, its impact on inflation past and present… and the inflation yet to come
26. Dobija, M., 2011. Abstract nature of money and the modern equation of exchange. *Modern Economy*, *2*(02), pp.142-152.
27. E. Marchi, G. Ferroni, F. Eyben, L. Gabrielli, S. Squartini, and B. Schuller, "Multi-resolution linear prediction based features for audio onset detection with bidirectional LSTM neural networks," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2014, pp. 2164–2168.
28. Elliott, Larry. "Greek debt crisis: Where did it all go wrong?." The Guardian (2019).
29. Eurostat, ec.europa.eu
30. F. A. Gers and J. Schmidhuber, "Recurrent nets that time and count," in Proc. IEEE-INNS-ENNS Int. Joint Conf. Neural Netw. (IJCNN), vol. 3. Jul. 2000, pp. 189–194.
31. F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," in Proc. 9th Int. Conf. Artif. Neural Netw. (ICANN), vol. 2. Sep. 1999, pp. 850–855.

32. Fid Backhouse and others, "hyperinflation in the Weimar Republic". Encyclopedia Britannica, 6 May. 2022, https://www.britannica.com/event/hyperinflation-in-the-Weimar-Republic.
33. Financial Times. (2022, March 1). Russian invasion of Ukraine risks wider economic shockwaves. Financial Times.
34. Fischer, David Hackett. *The great wave: Price revolutions and the rhythm of history*. Oxford University Press, USA, 1999.
35. Fisher, I., 1912. Will the Present Upward Trend of World Prices Continue?. *The American Economic Review*, *2*(3), pp.531-558.
36. Fisher, Irving. "" The Equation of Exchange" for 1911, and Forecast." *The American Economic Review* 2.2 (1912): 302-319.
37. Frain, J., 1995. “Econometrics and Truth”, *Central Bank of Ireland Technical Paper 2/RT/95.*
38. Frain, John. "Lecture notes on univariate time series analysis and box jenkins forecasting." *Economic Analysis, Research and Publications* (1992): 189-194.
39. Gatch, Loren. 1996. “To Redeem Metal with Paper: David Hume’s Philosophy of Money.” Hume Studies 22 (April): 169–92.
40. Geman S, Bienenstock E and Doursat T (1992). Neural networks and the bias=variance dilemma, Neural Comput 5: 1±58.
41. Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn and TensorFlow. O'Reilly Media.
42. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
43. Grauwe, Paul De, and Magdalena Polan. "Is inflation always and everywhere a monetary phenomenon?." *Scandinavian Journal of economics* 107.2 (2005): 239-259.
44. Greene, William H., and David A. Hensher. "A latent class model for discrete choice analysis: contrasts with mixed logit." *Transportation Research Part B: Methodological* 37.8 (2003): 681-698.
45. H. Sak, A. Senior, and F. Beaufays, “Long short-term memory recurrent neural network architectures for large scale acoustic modeling,” in Proc. Annu. Conf. Int. Speech Commun. Assoc. (INTERSPEECH), 2014, pp. 338–342. [Online]. Available: http://193.6.4.39/~czap/letoltes/ IS14/IS2014/PDF/AUTHOR/IS141304.PDF
46. Hall, Aaron Smalter. "Machine learning approaches to macroeconomic forecasting." *The Federal Reserve Bank of Kansas City Economic Review* 103.63 (2018)
47. Hamilton, J. D. 1994. *Time series analysis*, volume 2. Princeton university press Princeton.
48. Hashem S (1997). Optimal linear combination of neural networks. Neural Networks, 10: 599±614.

49. Hodge, V., & Austin, J. (2004). A survey of outlier detection methodologies. Artificial Intelligence Review, 22(2), 85-126.
50. Hume, D., 1752. *Political Discourses by David Hume Esq*. R. Fleming, For A. Kincaid and A. Donaldson.
51. Hutchison, Terence W. 1988. Before Adam Smith: The Emergence of Political Economy, 1662–1776. Oxford: Blackwell.
52. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice (2nd ed.). OTexts. Retrieved from https://otexts.com/fpp2/
53. International monetary fund, imf.org
54. J. Donahue et al. (Nov. 2014). "Long-term recurrent convolutional networks for visual recognition and description." [Online]. Available: http://arxiv.org/abs/1411.4389
55. J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, "DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. LDC93S1, 1993.
56. Jackson, Andrew, Ben Dyson, and Graham Hodgson. "The positive money proposal." *Positive Money: London* (2013).
57. Ji, Wenying. (2018). Simulation_Based Analytics for Fabrication Quality-Associated Decision support.
58. Jiang, J., Fu, F., Yang, T., Cui, B., 2018, Sketchml: Accelerating distributed machine learning with data sketches. In Proceedings of the 2018 International Conference on Management of Data (pp. 1269-1284).
59. Jones, A. H. M. "Inflation under the Roman Empire." *The Economic History Review*, vol. 5, no. 3, 1953, pp. 293–318. *JSTOR*, https://doi.org/10.2307/2591810.
60. Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M., & Tang, P. T. (2016). On large-batch training for deep learning: Generalization gap and sharp minima. In International Conference on Learning Representations (ICLR).
61. Keynes, John Maynard, Donald Edward Moggridge, and Elizabeth S. Johnson. *The Collected Writings of John Maynard Keynes*. Vol. 30. London: Macmillan, 1971.
62. Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.
63. Klaus Greff, Rupesh K. Srivastava, Jan Koutn´ık, Bas R. Steunebrink, Ju¨rgen Schmidhuber , "LSTM: A Search Space Odyssey", 2017
64. Lindemann, Benjamin, et al. "A survey on anomaly detection for technical systems using LSTM networks." *Computers in Industry*131 (2021): 103498.

65. Litterman, R., 1986. “Forecasting with Bayesian Vector Autoregressions - Five Years of Experience”, *Journal of Business and Economic Statistics*, January, No. 1, Vol. 4, pp. 25-38.
66. Little, R. J., & Rubin, D. B. (2014). Statistical analysis with missing data (2nd ed.). John Wiley & Sons.
67. Lohmann, L., 2005. Malthusianism and the Terror of Scarcity. *Making threats: Biofears and environmental anxieties*, pp.81-98.
68. Loshchilov, I., & Hutter, F. (2018). Decoupled weight decay regularization. In International Conference on Learning Representations (ICLR).
69. M.-T. Luong, I. Sutskever, Q. V. Le, O. Vinyals, and W. Zaremba. (2014). “Addressing the rare word problem in neural machine translation.” [Online]. Available: http://arxiv.org/abs/1410.8206
70. Makinen, Gail E. “The Greek Hyperinflation and Stabilization of 1943-1946.” *The Journal of Economic History*, vol. 46, no. 3, 1986, pp. 795–805. *JSTOR*, http://www.jstor.org/stable/2121485.
71. Makridakis S (1989). Why combining works? Int J Forecast 5: 601±603.
72. Malthus, Thomas Robert. *An Essay on the Principle of Population.*. 1872.
73. Mcleay, M., Radia, A., Ryland, T., “Money creation in the modern Economy, Bank of England Quarterly Bulletin, 2014 Q1
74. Meng, Y., Yang, N., Qian, Z., & Zhang, G. (2020). What makes an online review more helpful: an interpretation framework using XGBoost and SHAP values. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(3), 466-490.
75. Meyler, Aidan and Kenny, Geoff and Quin, Terry, “Forecasting irish inflation using ARIMA models”, Central Bank and Financial services of Ireland, 1998.
76. Mill, J.S., 1848. Principles of political economy with some of their applications. *Social Philosophy*, *1*, pp.1-329.
77. Mills, T. C., & Patterson, K. (2011). The characteristics and causes of seasonal inflation in the UK. Applied Economics, 43(24), 3427-3439.
78. Mishkin, F. S. (2007). The Economics of Money, Banking, and Financial Markets. Pearson Education. (Chapter 14: Money, Interest Rates, and Exchange Rates)
79. Nakamura, Emi. "Inflation forecasting using a neural network." *Economics Letters* 86.3 (2005): 373-378.
80. Oren Barkan, Jonathan Benchimol, Itamar Caspi, Eliya Cohen, Allon Hammer, Noam Koenigstein. Forecasting CPI Inflation components with Hierarchical Recurrent Neural Networks, International Journal of Forecasting, Volume 39, Issue 3, 2023.

81. P. J. Werbos, “Generalization of backpropagation with application to a recurrent gas market model,” Neural Netw., vol. 1, no. 4, pp. 339–356, 1988.
82. Paranhos, Livia. "Predicting inflation with neural networks." *arXiv preprint arXiv:2104.03757* (2021).
83. Pascanu, Razvan, Tomas Mikolov, and Yoshua Bengio. "On the difficulty of training recurrent neural networks." *International conference on machine learning*. Pmlr, 2013.
84. Phillips, Alban W. "The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861-1957." *economica* 25.100 (1958): 283-299.
85. R. J. Williams, “Complexity of exact gradient computation algorithms for recurrent neural networks,” College Comput. Sci., Northeastern Univ., Boston, MA, USA, Tech. Rep. NU-CCS-89-27, 1989.
86. Ross, E.B., 2003. Malthusianism, capitalist agriculture, and the fate of peasants in the making of the modern world food system. *Review of Radical Political Economics*, *35*(4), pp.437-461.
87. Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. *Learning internal representations by error propagation*. California Univ San Diego La Jolla Inst for Cognitive Science, 1985.
88. S. Chatterjee, A. Dash, and S. Bandopadhyay, “Ensemble support vector machine algorithm for reliability estimation of a mining machine,” Quality and Reliability Engineering International, 2014.
89. S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Dept. Fakultät für Informatik, Tech. Univ. Munich, Munich, Germany, Tech. Rep. FKI-207-95, Aug. 1995. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.51.3117
90. S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997. [Online]. Available: http://www.bioinf.jku.at/publications/older/2604.pdf
91. S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber, “Gradient flow in recurrent nets: The difficulty of learning long-term dependencies,” in A Field Guide to Dynamical Recurrent Networks, S. C. Kremer and J. F. Kolen, Eds. Piscataway, NJ, USA: IEEE Press, 2001.
92. S. K. Sønderby and O. Winther. (Dec. 2014). “Protein secondary structure prediction with long short term memory networks.” [Online]. Available: http://arxiv.org/abs/1412.7828
93. Schabas, M. and Wennerlind, C. eds., 2008. *David Hume's Political Economy*. Routledge.
94. Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. Psychological Methods, 7(2), 147-177.

95. Stewart, John B. 1963. The Moral and Political Philosophy of David Hume. Westport, CT: Greenwood.

96. Stock, J.H. and Watson, M.W., 2008. Phillips curve inflation forecasts.

97. Stockton, D., and J. Glassman, 1987. “An Evaluation of the Forecast Performance of Alternative Models of Inflation”, *Review of Economics and Statistics*, Vol. 69, No. 1, February, pp. 108-117.

98. Stracca, L., Terlizzese, D., & Bonci, R. (2013). Capacity utilization and inflation in the euro area. Journal of Policy Modeling, 35(2), 181-197.

99. T. G. Dietterich, “Ensemble methods in machine learning,” in Multiple classifier systems. Springer, 2000, pp. 1–15.

100. The American Heritage Dictionary, 2022

101. Totonchi, Jalil. "Macroeconomic theories of inflation." *International conference on economics and finance research*. Vol. 4. No. 1. 2011.

102. V. Pham, T. Bluche, C. Kermorvant, and J. Louradour. (Nov. 2013). “Dropout improves recurrent neural networks for handwriting recognition.” [Online]. Available: http://arxiv.org/abs/1312.4569 [5] P. Doetsch, M. Kozielski, and H. Ney, “Fast and robust training of recurrent neural networks for offline handwriting recognition,” in Proc. 14th Int. Conf. Frontiers Handwriting Recognit., 2014, pp. 279–284. [Online]. Available: http://people.sabanciuniv.edu/ berrin/cs581/Papers/icfhr2014/data/4334a279.pdf

103. Vickers, Douglas. 1959. Studies in the Theory of Money, 1690–1776. New York: Kelley.

104. W. Zaremba, I. Sutskever, and O. Vinyals. (Sep. 2014). “Recurrent neural network regularization.” [Online]. Available: http://arxiv.org/ abs/1409.2329

105. Wennerlind, C., 2005. David Hume’s monetary theory revisited: was he really a quantity theorist and an inflationist?. *Journal of political economy*, *113*(1), pp.223-237.

106. Williams, R. J. and Zipser, R. A., A learning algorithm for continually training neural networks. Neural computation, vol. 1, 1989, pp. 270–280.

107. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data mining: Practical machine learning tools and techniques (4th ed.). Morgan Kaufmann.

108. Xia, F., Liu, J., Nie, H., Fu, Y., Wan, L., & Kong, X. (2019). Random walks: A review of algorithms and applications. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *4*(2), 95-107.

109. Xia, Feng, et al. "Random walks: A review of algorithms and applications." *IEEE Transactions on Emerging Topics in Computational Intelligence* 4.2 (2019): 95-107.

110. Y. Fan, Y. Qian, F. Xie, and F. K. Soong, “TTS synthesis with bidirectional LSTM based recurrent neural networks,” in Proc. Interspeech, 2014, pp. 1964–1968.

111. Zhang X and Hutchinson J (1994). Simple architectures on fast machines: practical issues in nonlinear time series prediction. In: Weigend AS and Gershenfeld NA (eds). Time Series Prediction: Forecasting the Future and Understanding the Past. Addition-Wesley: Reading, MA, pp 219±241.