Prediction of the Consumer Price Index of Norway

Joana Simões
Departement of Informatics and Engineering
University of Coimbra
Coimbra, Portugal
joanasimoes@student.dei.uc.pt

***Abstract*—This paper explores the Norwegian Consumer Price Index (CPI) from 1990 until 2022. The data was divided into train and testing, using the 80-20 rule. First, it was carried out a global analysis of the time series (TS) finding that the original TS wasn't stationary, and, based on that, a first-order differencing and seasonal differencing was applied to remove trend and seasonality. The stationarity of the resulted TS was also checked with the Dickey-Fuller (DF) test, confirming its stationarity at a significance level of 95%. After that, several models were evaluated to discover the best forecasting models for the data ahead. Multivariate models were also tested, using the CPI index of the European Union and the United States as explanatory TS. In the end, based on the models' performance, it was concluded that the best forecasting model was SARIMAX achieving an R Squared of 0.989 and MAPE of 0.004.**

***Index Terms*—CPI, Time Series, Forecasting**

I. Introduction

The Consumer Price Index (CPI) is one of the most important macroeconomic indicators of a country. The CPI is an indicator used to calculate a nation's inflation rate, since it monitors monthly fluctuations in the value of a nation's currency as well as the cost of a variety of goods and services, hence the great importance of CPI forecasting [1]. This paper explores the Norwegian CPI index, first by analysing its past values, then it tries to forecast its future values.

The data contains the monthly CPI from 1950, but for the purpose of this work, the project focuses on the data from 1990 until August 2022. Later in this project, CPI values from other countries will be used to predict the Norwegian CPI. In total, it were used 329 samples, 313 for training the predictive models and for the exploration and 79 for assessing the models' performance, equivalent to an 80-20 split of the data.

Forecasting CPI has a great importance for a country. Álvarez-Díaz et Al. studied the influence of non-linear, Neural Networks and Genetic Programming, models against linear ones, Autoregressive (AR) and SARIMA, in forecasting the US CPI Indicator [6]. Their work concluded that use of non-linear models doesn't lead to better results, especially since the best model they obtained was SARIMA, although it did not achieve significantly better results.

In 2020, Riofrío et Al. carried out a comparative study of different predictive models to forecasting Ecuador's CPI from January 2005 to June 2019 [4]. Some models tested were Support Vector Regression (SVR), Neural Networks with Long Short-Term Memory layers, Seasonal Autoregressive Integrated Moving Average (SARIMA), Exponential Smoothing and Facebook's Prophet. The models were also optimized using grid-search to find the best hyperparameters. Using the last 12 months to evaluate the models, their study concluded that the model with the lowest Mean Absolute Percentage Error (MAPE) was the SVR, followed by the LSTM and SARIMA.

A similar approach was followed by Shinkarenko et Al. [5], when they predicted Ukraine's CPI from 2010 until 2020. In their work, they first analysed the data, based on trend, seasonality and them used the Box-Jenkins (ARIMA) and Exponential Smoothing models to predict the CPI of October and November 2020.

Nguyen et Al. [3] used more modern approaches to predict the monthly CPI index from the United States from 2017 until 2022. First, they divided into 80-20 for training and testing, then they tested several models, including multivariate ones, such as Multivariate Linear Regression (MLR), SVR, Autoregressive Distributed Lad (ARDL), and Multivariate Adaptive Regression splines (MARS) for forecasting. The CPI variable also relied on additional independent variables, such as the crude oil price, the global gold price, and the federal fund effective rate. In their experiments, they concluded that the SRV model achieved better results on the train data, but the MARS model was the winner on the test data.

In short, different approaches have been tested by many authors to predict a country's CPI. The most suitable technique seems to differ from dataset to dataset, and there is no general technique that always produces good results. It should also be noted that multivariate techniques seem to provide good results for predicting a dependent variable, but they require more data, which is not always possible.

This project aims to test some of these approaches, and find the ones that better forecast the Norwegian CPI Index. The techniques used along the project include classical forecasting models such as ARIMA and SARIMA, multivariate models such as VARMAX and SARIMAX and Machine Learning models in the univariate and multivariate context.

The remainder of the paper is organized as follows. Section II presents a description of the methods developed throughout the project. In Section III, the results obtained by the different techniques tested are detailed. Section IV summarizes the main findings of this paper. Finally, Section V presents the conclusions of these papers and the future work that needs to be done.

II. METHODS

This section presents all the methods developed along the project. It includes a first analysis and preprocessing of the TS in order to make it stationary. Subsequently, it presents the methodology employed to evaluate diverse forecasting models.

A. Preprocessing and analysis of the Time Series

The initial dataset had a lot of data that wasn't relevant to the problem. As a result, it needed to be decluttered. Additionally, the date and time information wasn't in the proper format to be used. As a result, after the dataset had been tidied up, it only had a first column for the date and time, then a column for the CPI index of each country.

The data was then divided into training a testing data, using the rule of 80% for training and 20% for test. With the data divided, the training part was used for characterization of the TS, modelling and training. The test data was saved for later evaluation of the forecast models' performance.

With the selected TS, the first task was to do a brief analysis of it. It was checked if it had missing values, confirming that it had not. Following that, the stationarity of the TS was examined. First, it was analysed through the visualization of the time series (TS), and it was observed that the TS had a clearly increasing trend and some seasonality (see Figure 1). This concluded that the TS is not stationary. In addition to the visual analysis of the stationarity, the Dickey-Fuller (DF) test was applied to confirm the results through empirical observation. Using a significance level of 95%, the DF statistical test assumed the hypothesis:

- H0: The TS is non-stationary;
- H1: The TS is stationary;

The test (see Table I) verified that the TS is non-stationary. This same statistical test was used throughout the project to determine whether the TS was stationary, always using the same hypothesis and significance level As the original TS is non-stationary, it needs to be transformed in order to become it. To achieve this, it was employed a decomposition model. These models are based on the idea that the TS have trend and cyclic parts, and they divide the TS in its trend, seasonal and erroneous (residuals) components.

To determine the appropriate decomposing model for the problem ahead, it was analysed the magnitude of the values. The two main choices, as the values are distant from zero, are the additive or the multiplicative model. When the order of magnitude of values does not change throughout time, it is used an additive model, otherwise a multiplicative model should be applied. As it can be observed by the Figure 1, the order of magnitude doesn't change over time, so it was applied an additive model.

In the additive model, the TS trend is represented by $x(n) = tr(n) + sn(n) + e(n)$, where $x(n)$ is the time series, $tr(n)$ the trend component, $sn(n)$ the seasonal component, and $e(n)$ the erratic component.

First, The trend was removed by fitting a model for the 1$^{\text{st}}$, 2$^{\text{nd}}$ and 3$^{\text{rd}}$ order trends. After select the most appropriate, the trend adjusted TS was calculated by subtracting the trend to the original TS. The trend adjusted TS was initially subjected to a frequency analysis. With this information, a low-pass filter was applied to the TS while taking the most crucial frequencies into consideration to calculate the seasonality using a Butterworth low-pass filter. Once the seasonal component had been identified, the erratic component was determined by removing the seasonality from the trend adjusted TS. The order of the fitting model and filter was adjusted according to results achieved by the Dickey-Fuller test.

Another approach to reach a stationarity is by differencing. Initially, the trend is removed by differentiation. After that, the seasonality differencing is applied with a period of 12, equal to the number of samples per year. The orders used were selected based on the train data.

These two approaches were evaluated and the one that achieved better results in the DF test, was used in the next steps of the process. The Autocorrelation Sequence (ACS) of the two approaches was also tested. However, it was not applied merely to visually observe how the time series behaved. In the next stage of the project, this will be important in choosing the type of model that represents the stationary TS.

B. Forecasting Models

Several models were tested to solve the forecasting problem. These model vary in their internal architecture, some use only univariate models, others can use multivariate models. Some models are linear, others are non-linear.

In this section, the TS was analysed to determine the most appropriate hyperparameters for the models. At the same time, the train data was separated into train and validation, and a grid-search was performed to find the best hyperparameters, using the Box-Jenkins Method. The validation data was constructed by selecting the final 15% of the train data. The selection of the best hyperparameters was made using the Mean Absolute Percentage Error (MAPE) of the predictions on the validation data.

After selecting the best hyperparameters for the models, they were tested using different forecast horizons. It was decided to test four different forecast horizons, 1, 3, 6 and 12 months. However, in some models, it was forecasted the entire test data instead of using the iterative process, since that in some models, Exponential Smoothing and in ML models, it was not possible to use a different forecasting horizon

Initially, it were tested linear models for the stationary data, previously transformed using differencing. A first analysis of the ACS and Partial ACS of the stationary time series was used to select the hyperparameters p and q for the Moving Averages (MA) and Autoregressive (AR) part of the linear model, respectively. These parameters were then confirmed by a grid-search, indicating that an Autoregressive Moving Average (ARMA) should be used, instead of only an MA or AR model.

Next, it were evaluated models for the original non-stationary data. The first model was a Seasonal Integrated ARMA (SARIMA) since the data presents a seasonality component. In order to find its best hyperparameters, the TS was analysed. As it presented a trend and seasonality each 12 month, a first order differencing to remove the trend and seasonality was applied. With the stationary TS and a lag inferior to the period (12), the ACS and PACS were observed to select the most adequate hyperparameters q and p, respectively. The values were selected based on the order that fall out the unit root, meaning that it were significant different from zero. To select the seasonal order of Q and P, the ACS and PACS for the lags 12, 24, and 36 were examined. Once again, the best hyperparameters were confirmed with a grid-search using the train and validation data.

The other model evaluated was Triple Exponential Smoothing with an additive model. For this model, it was needed to find the most adequate values for the model level (α), for the trend (γ) and seasonality (β). A grid-based search was employed again for this purpose.

Until now, all the models applied were univariate. Nonetheless, this work also tested multivariate models. Since the goal of this project is to forecast the CPI index of Norway, the explanatory TS used as complementary should be of countries or organizations that have major relationships with Norway. After an analysis of the main trade partners of Norway, it was found that most of them were members of the European Union(EU) [2]. Therefore, the CPI index of the EU was calculated by averaging the CPI index of its members in 1990, the initial year of the TS. Another major parter of the Norway and that also influences the world economy is the United States (US), so its CPI Index was also included as explanatory TS for the forecasting. After an initial analysis of the correlations between these three CPI Index, it was decided to use a Single equation model, Seasonal Autoregressive Integrate Moving Average model with Exogenous inputs (SARIMAX), since it is the lag 0 that has higher correlation. Another model applied was the Vector Autoregressive Moving Average with exogenous inputs (VARMAX). The second one needs stationary time series as inputs, so the TSs were transformed using the differencing described previously.

To encompass the entire spectrum of forecasting techniques, we also evaluated ML-based techniques for this project. Similar to [4], the project applied Recurrent Neural Networks (NN), specially Long Short-Term Memory (LSTM) since it takes into account long and short time dependencies and Gated Recurrent Units (GRU). Both were evaluated in a univariate and multivariate context. Since NN achieve better results if the data is normalized, in a range 0 to 1, the data was normalized using a Min Max Scaler before it was given to the NN models. The ML models need a different input data. It needs the data in windows, with a shape *number samples x size of the input sequence*. These inputs are created by sliding a window (input sequence) over the TS, using the previous X lags to predict each. It was decided to use an input sequence (X) of 12 lags, meaning that each point is predicted based on the previous year. Additionally, it was added an early stopping mechanism to the ML models so that it stops the train process if the validation loss stops decreasing. Several hyperparameters were tested manually to tune the ML models, and the bests models found are described in the next section.

To summarize, a variety of distinct forecasting models were evaluated in order to determine the most suitable one for the forthcoming task.

III. Results

The Section present the main results found in this work. First, it presents an overall analysis of the Norwegian CPI value and work made to convert the time series into stationary. Finally, the results of the different forecasting models are shown.

A. Analysis of the Time Series

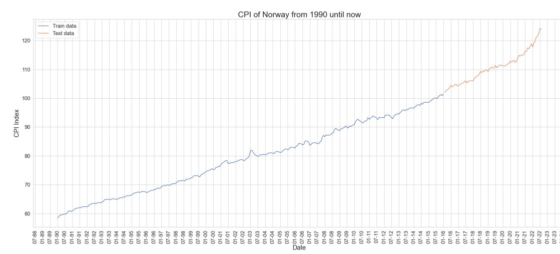

Fig. 1: CPI Index of the Norway from 1990 until 2022

In a first analysis of the TS (see Figure 1), it can be observed an increasing trend and patterns repeated every 6 months, and every year. In addition, it can be seen that the order of magnitude of the values does not seem to vary over time, so an additive decomposition model was used. After careful examination of the Time Series, it seems that the trend shifts from a linear trend to a larger order trend after 2020. This shift can be explained by the COVID-19 era, which altered the world and caused huge economic changes. Since this shift appears only on the test data, in the future, the models could have difficulties in correctly forecast the test data.

B. Basic Transformations on the TS

In Section II it was concluded that one possible approach to transform the series into stationary was using an additive model to decompose it. Therefore, to transform the TS in stationary, the first step was to remove the trend. This was achieved using model-fitting, where the 1st, 2nd and 3rd order trends were tested. According to the results, the 1st order trend was adequate for the problem, since it is a simpler model and the differences to the other orders are insignificant.

After removing the trend by subtraction from the original series, it can be seen more clearly the seasonal patterns of the TS. The patterns seem to repeat every six months, every year, and approximately every five years. The results were further strengthened by the Discrete Fourier Transform (DFT), where it can see the peaks at T=0.5 (6 months), at T=1 (every year) and at T=5 (every 5 years). It was also found that the most important seasonal component is 2/3 months. After testing several cut-off points for the Butterworth filter, it was decided

	Initial TS	1st Approach	2nd Approach	2nd Approach on Test Data
ADF value	1.00	-17.22	-7.24	-2.99
p-value	0.99	0.0	0.0	0.03
1%	-3.45	-3.45	-3.45	-3.45
5%	-2.87	-2.87	-2.87	-2.87
10%	-2.57	-2.57	-2.57	-2.57

TABLE I: DF Statistical Test. 1st approach: Trend as seasonal adjusted by model-fitting and filtering; 2nd approach: Trend and seasonal adjusted by differencing

to use a cut-off point of 2.2, to pick only the most important frequencies and with an N=5.

With the trend adjusted series and the seasonal component, it was calculated the erratic component was calculated by once again subtracting the seasonal component to the trend adjusted series.

A second approach to transform the TS in stationary was using differencing. The results in Figure 2 show that there still seems to be some seasonality. In order to confirm the results of the two approaches empirically, the DF test was applied.

Table I presents the p-values obtained by the approaches. Although, the results present in Table I indicate that the TS transformation into stationary using model fitting and filtering reached better results, in the next stages of the project it will be used differencing to transform the TS. This decision is motivated by the fact that it is easier to invert the transformation, at the same time, by the fact that it is the mechanism used internally by many forecasting models.

The ADF test on the test data tell that although the TS is stationary for a significance level of 95%, it is more close to the critical values, i.e, it is in the limit of not being stationary, even failing if it was used a significance level of 99%.

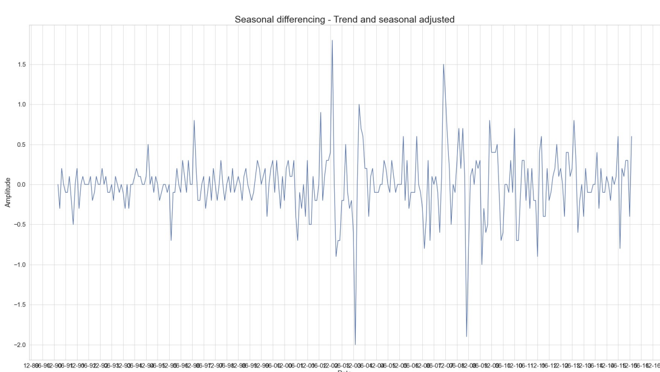

Fig. 2: Trend and seasonal TS adjusted by differencing

s

A further analysis of the stationary TS was carried out. Figure 3 presents the ACS values of the stationary TS using a significance level of 95%. The ACS shows a rapid decrease, reaching a value close to zero, which indicates that the series is stationary. The peaks at the lags multiple of the period in Figure 3 and Figure 4 could indicate that some seasonality is

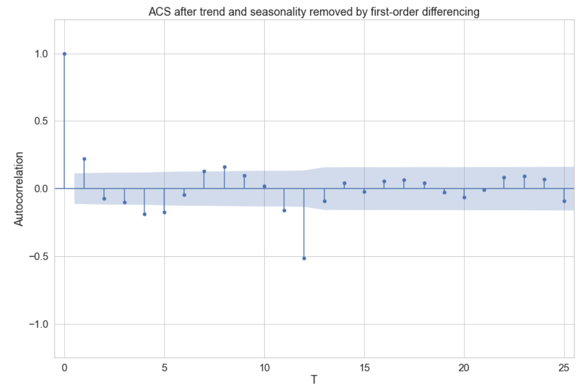

Fig. 3: Autocorrelation Sequence of the stationary Time series

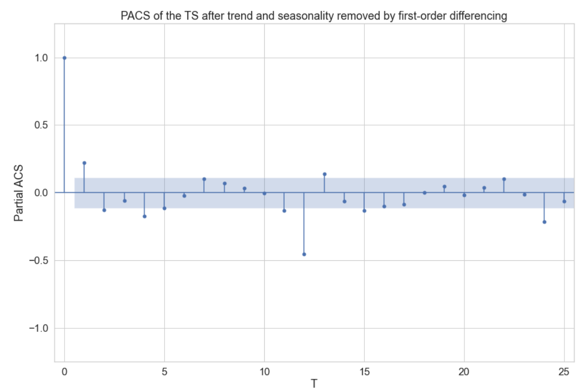

Fig. 4: Partial Autocorrelation Sequence of the stationary Time series

still present, but since the ADF test indicated that the TS is stationary, no higher order differencing was carried out.

Other TS were used to as exogenous variables and given to the models along with the original TS. It can be seen that they all present a similar pattern, having an increasing tendency and some seasonality (see Figure 5). Figure 6 presents the cross-correlation between the variable of interest and the explanatory TS. It is observed that there are few unidirectional relations between our TS of interest and the explanatory TS, except at lag T=0, i.e, relation in the exact point. Other relations appear when T=12, meaning that the CPI Index on the previous year for the EU and US influence the TS at each point.

C. Models

In order to find the best forecasting models, several architectures were tested. The architectures include ARIMA, SARIMA, Triple Exponential Smoothing, and SARIMAX, VARMAX and Machine Learning (ML) models for multivariate predictions.

Table II presents the results achieved by the different models tested on its best forecasting horizon. It can be concluded

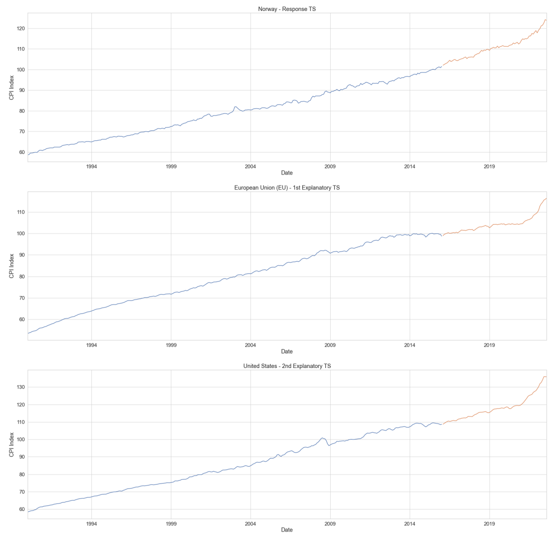

Fig. 5: Distribution the interest TS and the CPI Index of EU and US, used as explanatory TS, separated into train and test data

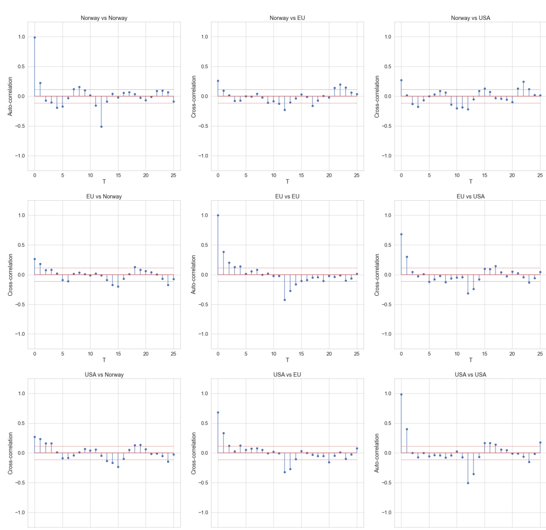

Fig. 6: Cross-correlation between the interest TS, Norway CPI index and the explanatory TS, from the United States and European Union

Model	H	MAPE	MAE	MSE	RMSE	R2
ARIMA(2, 4)	6	0.010	1.160	2.449	1.565	0.916
SARIMA(4, 1, 2)x(3, 1, 0, 12)	3	0.005	0.553	0.659	0.811	0.977
Exponential Smoothing	—	0.024	2.666	10.212	3.196	0.648
SARIMAX(1, 1, 0)x(1, 1, 1,12)	3	0.004	0.442	0.328	0.573	0.989
VARMAX(1, 2)	12	0.039	4.459	36.533	6.044	-0.259
ML model 1	1	0.021	2.350	8.399	2.898	0.710
ML model 2	1	0.032	3.599	17.075	4.132	0.411
ML model 3	1	0.057	6.560	69.528	8.338	-1.397
ML model 5	1	0.119	13.575	292.058	17.090	-9.067

TABLE II: Results achieved by the TS models for the best forecast horizon

ML Model	RNN	Stationary Data
1	GRU	False
2	GRU	False
3	GRU	True
4	GRU	True

TABLE III: Hyperparameters used on the ML Models. Note: All the models were trained in 100 epochs with a RMSprop optimiser and a using the MSE as loss and a batch size of 1. It was also added an early stopping mechanism to avoid overfitting.

that the SARIMAX model has the minimum error, followed by the SARIMA model. Both of them, are good to handle the seasonal patterns of the TS, and reach the maximum performance with a forecasting horizon of 3 months. Then, the ARIMA model is the third best forecasting models. The rest of the models have more difficulty in the predictions, having higher prediction errors. The R2 values, which determines the proportion of variance in the dependent variable that can be explained by the independent variable, i.e., indicates the accuracy of the estimation in relation to the original data, say that the SARIMAX, SARIMA and ARIMA models that the only that make accurate predictions since their R-Squared is above 0.9. The remaining models have lower accuracy in predicting the out-sample data.

From the predictions graph in Figure 7, the SARIMAX almost perfectly described the out-sample data. The reason

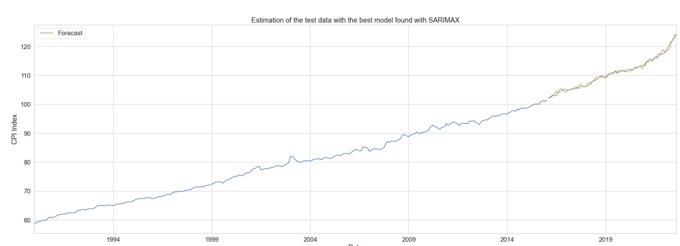

Fig. 7: Forecasts of the best SARIMAX model compared to the original TS

it was the best forecasting model may be related with the fact that it uses auxiliary TS to help the model in the predictions. These explanatory TS also have a similar CPI Index (see Figure 5), which aligns with the interest TS to make accurate predictions. This explanation can also be validated by the fact that the same architecture without exogenous variables also achieves good results, but slightly below.

Overall, the models that handle non-stationary data achieved better results when compared with the corresponding model that handles stationary data in the ML models. It could indicate that removing the trend and seasonality components of the TS, also removes important information that the models need to achieve good results.

Even though the models tested achieve satisfactory results, the out-sample data is not helping with the predictions. Figure 1 shows the in the test data, the TS changes from a linear trend to a higher polynomial trend, which means that the parameters learned during the train are not prepared to deal with this change, which could explain why some model have so poor results. The ADF test even indicates that for certain significance values, the transformed test TS is not stationary, i.e., the differencing applied based on an analysis of the train data is not the better for the test data. However, this problem is slightly controlled since the forecast horizon is only a couple of months, which means that the model updates are able to cope with the change in trend.

Only a few possible combinations of hyperparameters were tested for the ML models, many more were possible. However, since the o goal of the project is not finding a good configuration of hyperparameters for the neural networks, only some representative ones were evaluated to find if the NN architecture could be the solution for this problem.

IV. Discussion

The shift from a linear trend to a higher order polynomial in the test that impact the models negatively. As this change only appears on the out-sample data, the in-sample data fools the models. Even though, some models, specially the SARIMA and SARIMAX models, achieve satisfactory results on the out-sample data, with an R-square metric indicating that the models have good performance in explaining the test data.

The multivariate SARIMAX model turned out to be the best one. This model makes use of two additional TS, the US and EU CPI index, alongside the variable of interest. The findings suggest that in order to predict future values, more information is required than just the variable of interest.

Despite having lower performance, the second-best model, the SARIMA model, may be better in some applications rather than the SARIMAX. The justification is that the best model is multivariate, i.e. it requires additional TS to make good predictions in the future. However, this data has not always been measured and in the case of the CPI index of the countries, it is most likely that they are all calculated on a monthly basis, so it is not possible to access future data to predict Norway's data. Therefore, SARIMA could be a good alternative, as it only uses past data on the variable of interest to make predictions.

It is also worth noting that the results obtained by the different RNNs fall short of what is expected, given that these are state-of-the-art models, and that they are usually always ahead of the others. The explanation for their low performance may be related to the fact that these models generally require a lot of data, and in this context where they are applied, with monthly data since 1990, this may not be enough. Another reason is linked to the fact that there are many possible configurations for the RNNs and only a small percentage have been tested, and this may not be the best for the problem.

Another interesting finding was the lower performance of the Triple Exponential Smoothing. In the literature, this model was very used to predict the CPI Index of a country, but in this project, the results were much worse than other models.

V. Conclusion

This project consisted in exploring different models with the goal of finding the best forecasting model to predict the Norwegian CPI Index. The data used was from 1990 until 2022, divided in 80% for training and 20% to assess models accuracy.

The best model found used multiple TS, i.e, in addition to the variable of interest, Norway's CPI index, it also used the CPI indices of the European Union and the United States.

The next steps on this project, and as the best model was multivariate, may include trying the CPI Index of other countries as exogenous variables. For example, adding the China index, since it is the world's largest exporter, other the index of other economic partners of Norway. Another important task will be to test other configurations for the models, especially the ML models. Testing different optimizers, batch sizes, and other hyperparameters that change the RNN architecture.

References

[1] Consumer Price Index (CPI). *Corporate Finance Institute*.

[2] Foreign trade figures of Norway - International Trade Portal.

[3] Tien-Thinh Nguyen, Hong-Giang Nguyen, Jen-Yao Lee, Yu-Lin Wang, and Chien-Shu Tsai. The consumer price index prediction using machine learning approaches: Evidence from the United States. *Heliyon*, 9(10):e20730, October 2023.

[4] Juan Riofrío, Oscar Chang, E. J. Revelo-Fuelagán, and Diego H. Peluffo-Ordóñez. Forecasting the Consumer Price Index (CPI) of Ecuador: A Comparative Study of Predictive Models. *International Journal on Advanced Science, Engineering and Information Technology*, 10(3):1078, June 2020.

[5] Volodymyr Shinkarenko, Alexey Hostryk, Larysa Shynkarenko, and Leonid Dolinskyi. A forecasting the consumer price index using time series model. *SHS Web of Conferences*, 107:10002, 2021. Publisher: EDP Sciences.

[6] Marcos Álvarez Díaz and Rangan Gupta. Forecasting US consumer price index: does nonlinearity matter? *Applied Economics*, 48(46):4462–4475, October 2016. Publisher: Routledge _eprint: https://doi.org/10.1080/00036846.2016.1158922.