Is Prophet an effective comparator for building Inflation models?

21008019

MSc Data Science

Department of Computer Science and Creative Technologies, UWE

Bristol, UK

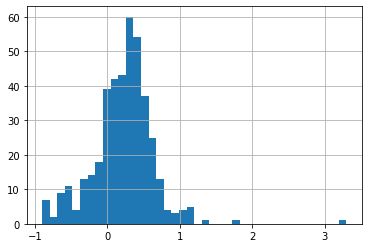


Figure 1 - % change CPI histogram

Keywords—SARIMA, ANN, Prophet, Consumer Price Index

# Introduction

In economics inflation is a time series measurement of the change of the prices of goods and services. It is often expressed as an index of producer price, house price or consumer price. The consumer price index gives the change in price of a weight average market basket of consumer goods purchased by a household. Economists have long been invested in building models for the forecast of inflation and over the last four decades, the Phillips curve has been widely regarded as among the most reliable and accurate prediction tools (Stock and Watson, 1999). Higher predictive accuracy yields increased benefit for monetary policy makers who must meet government targets. In the UK, the government’s inflation target is set at 2% (*Monetary policy*., n.d.).

Most of the literature regarding the subject does agree that multivariate methods, such as the Phillips curve outperform univariate methods in creating more accurate forecasts.

Univariate AR models are amongst the most favoured comparators upon which the performance of economic tools are evaluated (Forni *et al.*, 2003). This project will compare the performance of, Facebook’s, rather hubristically named, Prophet as a new comparative standard against the often-used ARIMA model.

# Exploratory Data Analysis

A picture containing line chart

Description automatically generatedThe time series collected from the Office of National Statistics (*CPI ANNUAL RATE 00: ALL ITEMS 2015=100 - Office for National Statistics*., n.d.). Includes the monthly percentage change of CPI in the UK from 02/88 to 03/22.

Figure 2 - % change CPI against Date

There are 410 values that range from -0.9 to 3.3 with a mean of 0.22. The data is relatively stationary over the sample period and is distributed normally with a standard deviation of 0.41.

There is also very little deviation from the seasonal pattern that repeats yearly.

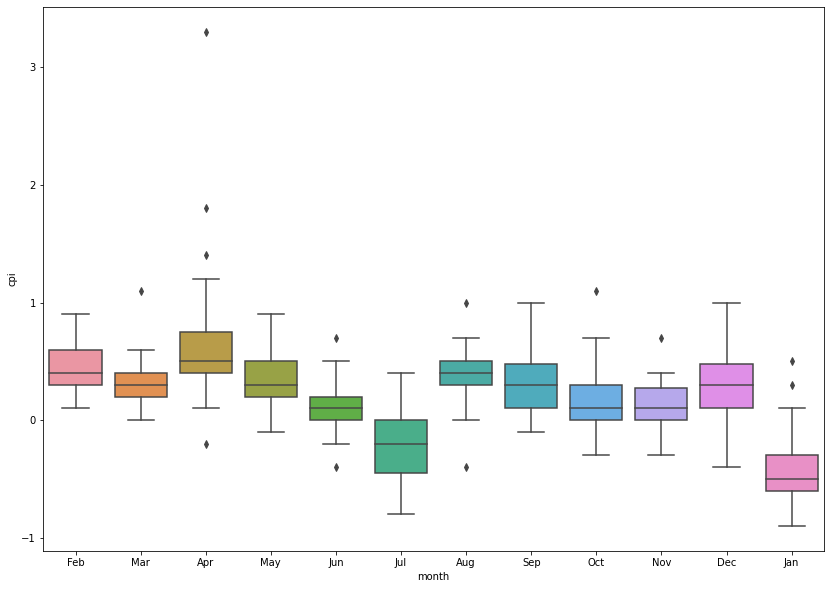


Figure 3 - Subset boxplots by month

# model Selection

An autoregressive integrated moving average model, ARIMA, is an analytical model used to model the behaviour of time series data and produce forecasts. They benefit from great flexibility in that many configurations can be implemented depending on the nature of the problem. In general, we describe ARIMA configurations in the form:

Where the values and are orders of the autoregressive and moving average polynomials respectively. The value references the order of difference components, necessary when the data is not stationary. Seasonality can also be captured with values , , , representing those seasonal components and *,* the period of the seasonal cycle.

The pmdarima package provides useful tools for estimation of the best model configuration. Autoarima uses a stepwise selection process to determine the best model by minimizing AIC, an estimator of predictive error. This process, while exceedingly convenient, can be costly as each model is takes some time to fit and that time increases with model complexity.

Further data analysis can help narrow down the search for the most effective model:

## Orders of AR and MA

By plotting the ACF, autocorrelation function and PACF, partial autocorrelation function. It is possible to determine a more specific range for the parameters of the autoarima.

Timeline

Description automatically generated

Figure 4 - ACF plot

Timeline

Description automatically generated

Figure 5 - PACF plot

To estimate the order of AR terms, , or the range of reasonably tested AR terms we look towards the PACF plot, which gives the additional correlation explained by each successive lagged term, and count the number of lags producing correlation outside the significance interval. Unfortunately here, the first lag falls within the significance interval, so is can be thought that there is no correlation.

Estimating MA terms, , is done in a similar fashion by inspection of the ACF plot, the correlation of the time series with a lagged version of itself. Here again, the first lag produces no significant correlation.

Unfortunately, it was not possible to derive a more specific search criteria for those terms. This is perhaps due to the low resolution of the dataset. Future research may consider the resampling of the dataset to create finer resolution before model selection.

## Seasonal Terms

The figures 4 and 5 do however confirm the existence of seasonality. Both show significant correlation at every 6 lags. It can then be inferred that the series is not totally random and there is some underlying pattern. This confirms the inference made during the initial exploratory phase and suggests the autoarima should also test for seasonal components with assumption that they will yield a more effective model.

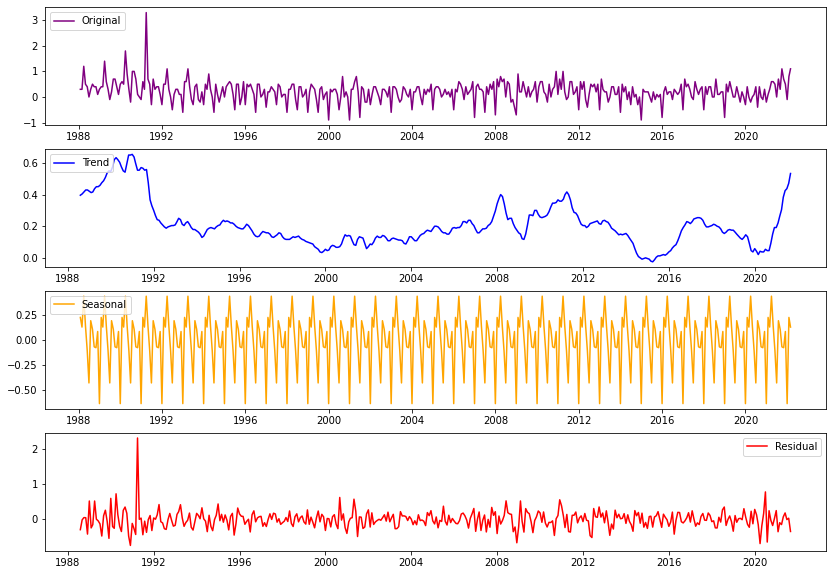


Figure 6 - Seasonal decomposition of time-series

Again, the existence of a seasonal component is confirmed. The decomposition shown in figure 6 captures the seasonal yearly pattern. Additional confidence in the model can be derived when residuals do not appear to maintain any systematic pattern as it falls in line with the assumptions of error of the regression (‘10.3 - Regression with Autoregressive Errors | STAT 462’, 2018).

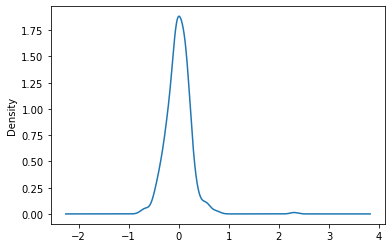


Figure 7 - Residuals of seasonal decomposition

## Order of Differencing

Having observed the movement of observations over the sample period. There is not much evidence to suggest there is any enduring time dependent component or “trend”. A Dickey-Fuller test can be applied to provide further evidence of stationarity(Mushtaq, 2011) .

Testing of the inflation data returned:

This exceeds the threshold required to reject the null hypothesis, that the data does not have a unit root. Therefore, some differencing factor must be taken into consideration.

## Autoarima

It’s somewhat ironic that this supplementary data analysis, which had been planned to reduce the set of reasonably tested models, not only failed to produce argument for a narrowing of the test field but created additional considerations. However, the insight provided may inform future research.

Fortunately, the dataset is of relatively few datapoints, meaning the individual model fits did not take an exorbitantly long time and the autoarima was able to arrive at an approximate solution within 83.478 seconds. The method returned the best model with configuration:

Note the period of seasonality, , was set at 12 as the data was collected monthly and there are 12 observations within every cycle.

## Prophet

Luckily the implementation process for Facebook’s Prophet algorithm was relatively simple. After the necessary formatting of the data the model was trained and tested. Despite it being perhaps too early, afore the presentation of any result, to make any comment of the benefits of Prophet. There is something to be said about the ease with which it’s implemented. Considering that a significant proportion of this report is dedicated to explaining the configuration of the baseline model while Prophet appears to arrive ready for training out of the box, does beg further questions. Further research regarding Facebook’s own motivation for the design of Prophet’s backend architecture may provide some interesting result. Are decision makers at Facebook interested in providing tools for researchers? Or is this investment another example of their proponency for quick results and short-term satisfactions. Is prophet merely a tool for plug and play, kebab-shop machine learning? (Hisham, some weeks ago)

# Results

A train test split of the dataset in the ratio 66:33 was performed. Train and test scores were given by computation of the RMSE and score.

|  |  |  |
| --- | --- | --- |
|  | ARIMA | Prophet |
| Train RMSE | 0.28 | 0.26 |
| Test RMSE | 0.28 | 0.32 |
| Score | 0.35 | 0.19 |

Table 1- Results

Prophet was outperformed by the baseline model in the test despite being able to produce a lower RMSE value while training. The both model’s test RMSE was 6.0% and 7.6% of the range of the timeseries respectively.

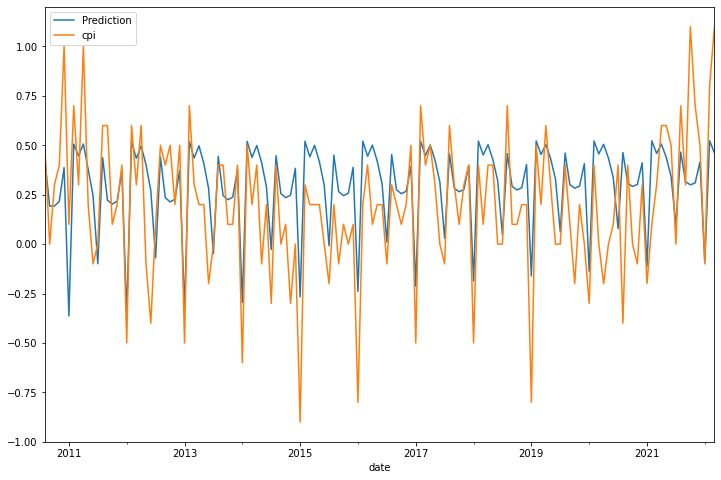


Figure 8 - ARIMA test predictions vs Actual Data

The ARIMA was able to capture much of the seasonality as is observed on the plot above. The blue prediction line seems to trace out the periodic transients in the test set.

A picture containing text, needle

Description automatically generated

Figure 9 - Prophet test prediction vs Actual Data

The Prophet test forecast was able to trace the periodic transients in the test data. Through the period 2012 to 2021 the model forecasted higher seasonal peaks that consistently overshot the actual data, whereas the ARIMA forecast tracked the change in local maximum more effectively.

# Discussion

Both models were only able to explain relatively small proportions of the total variance of the timeseries, with the ARIMA still outperforming Prophet significantly in this regard. Despite this both models seem to capture seasonality well, however the baseline seemed to capture more of the time- dependent trend. It is apparent that the underlying truth that drives the mechanism for the change in CPI is not best represented within the univariate domain, as was warned by the economists (Forni *et al.*, 2003). If we consider the context of the test period 2011 to 2022, where global economies were only just recovering from a period of financial downturn, there is some inferential reasoning for the relatively poor representation of the timeseries’ condition in both models. The ARIMA model benefits from it’s tunability in that it may be configured for a range of applications. This flexibility is lacking in Prophet, that provides fewer tunable parameters.

Future research may reframe the comparison by perhaps using seasonally adjusted variation of the dataset, which would reveal more of its underlying long term trend, of by increasing the resolution of the dataset. At its current resolution there was little correlation of the timeseries to the step shifted version of itself. Increasing its resolution may allow the models to capture more of the underlying mechanism for the CPI’s movement. However, this proposition does beg another question. Does this datatype lend itself towards finer resolution? Or can changes in the price of consumer goods be expressed in periods less then a month? In conclusion, we have provided little evidence to support Prophet’s viability as a comparator for inflation models in economic studies and the results suggest that ARIMA provides a more accurate baseline.

# Reference

‘10.3 - Regression with Autoregressive Errors | STAT 462’ (2018) *PennState Eberly College of Science*.Tutorial2018 [online]. Available from: https://online.stat.psu.edu/stat462/node/189/ [Accessed 18 May 2022].

*CPI ANNUAL RATE 00: ALL ITEMS 2015=100 - Office for National Statistics*. (no date) [online]. Available from: https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/d7g7/mm23 [Accessed 16 May 2022].

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2003) Do financial variables help forecasting inflation and real activity in the euro area? *Journal of Monetary Economics*. [online]. 50 (6), pp.1243–1255.

*Monetary policy*. (no date) [online]. Available from: https://www.bankofengland.co.uk/monetary-policy [Accessed 16 May 2022].

Mushtaq, R. (2011) TESTING TIME SERIES DATA FOR STATIONARITY. p.19.

Stock, J.H. and Watson, M.W. (1999) Forecasting inflation. *Journal of Monetary Economics*. [online]. 44 (2), pp.293–335.

Hisham. (some weeks ago) Kebab Shop Machine Learning. *Joke from class*. [in-person quote].