Consumer Price Index Analysis

Miguel Galvão
Department of Informatics Engineering
University of Coimbra
Coimbra, Portugal
galvao@student.dei.uc.pt

João Domingues
Department of Informatics Engineering
University of Coimbra
Coimbra, Portugal
domingues@student.dei.uc.pt

***Abstract*—In this paper we analyze the Consumer Price Index (CPI) in The Netherlands so as to assert the best methods to use to characterize a Time Series (TS) and predict its future values. We explore basic description techniques and transformations, model fitting, forecasting and even evaluate the relationship of the TS with that of another country.**

***Index Terms*—Time Series, TS, CPI, Model Fitting, Forecasting**

I. Introduction

Time Series Analysis is a field of Data Analysis in which data is gathered over time to be characterized and used to forecast future results. In this paper, we analyze the evolution of Consumer Price Index (CPI) of The Netherlands from January 1990 to August 2022, visible in Fig. 1. In the next section we present the methods used throughout this project.

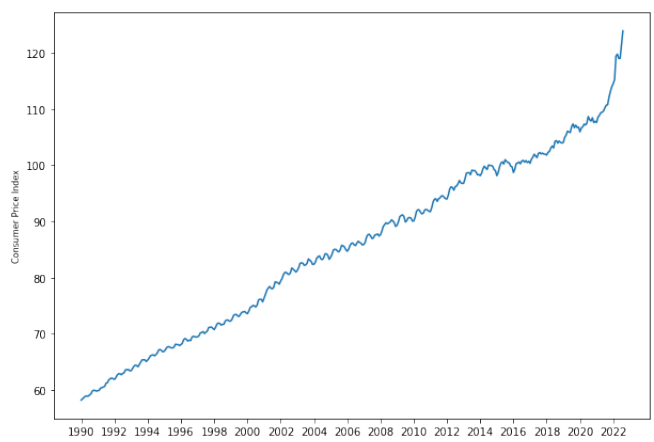

Fig. 1. Consumer Price Index (CPI) Time Series

II. Methods

By looking at the plot presented, we can see that the Time Series (TS) shows signs of increasing noise and pattern repetition over time. Because of that, we believe a multiplicative model is the most suited to describe its behaviour, as oppose to an additive or pseudo-additive one. The multiplicative model describes the TS, x(n), in the following way:

$$x(n) = tr(n)sn(n)e(n),$$

where tr(n), s(n) and e(n) represent, respectively the trend, seasonality and erratic components.

A. Trend Removal

The trend in a time series indicates its change in the long term, or in other words, the direction of the series over time. We will explore different techniques to try to characterize (and remove) this component from the TS.

1) Model Fitting: With this approach we attempt to fit a model to the TS curve in order to describe its evolution. This is done by calculating the coefficients of a polynomial function of a certain order so that it's as close as possible to the original TS. In Fig. 2 we can see the fitting of three different order polynomial functions and the resulting TS after removing the trend described by them. This removal is done by dividing the TS by the polynomial function, according to the multiplicative model. We tested a linear, quadratic and eighth order functions.

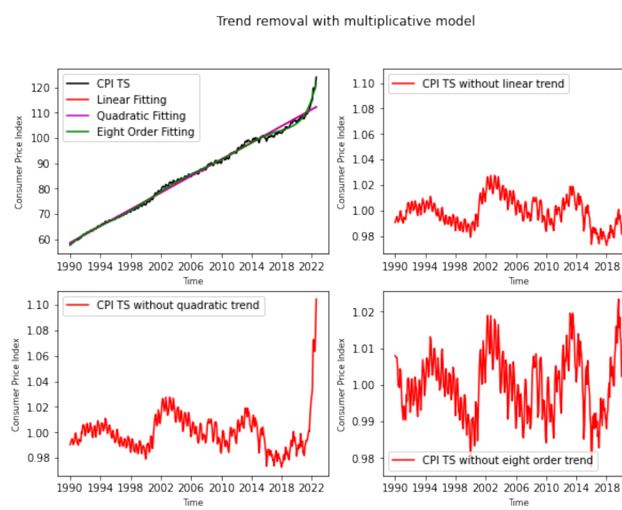

Fig. 2. Trend Removal with Model Fitting approach

2) Moving Average Filtering: As oppose to the previous approach where we attempted to characterize the trend in a global way, a filtering using a Moving Average (MA) applies a local smoothing technique, that better suits the TS.

As the name indicates, this method applies a moving filter that calculates the average of the values within according to a set of weights:

$$\hat{t}(n) = \frac{1}{\sum_{j=0}^{M} \omega_j} \sum_{k=n-\frac{M+1}{2}}^{n+\frac{M+1}{2}} \omega_k x(k)$$

where M is the size of the filter and ω the set of weights. We tested different values for M: 13 and 19. The resulting smoothed TS's are presented in Fig. 3.

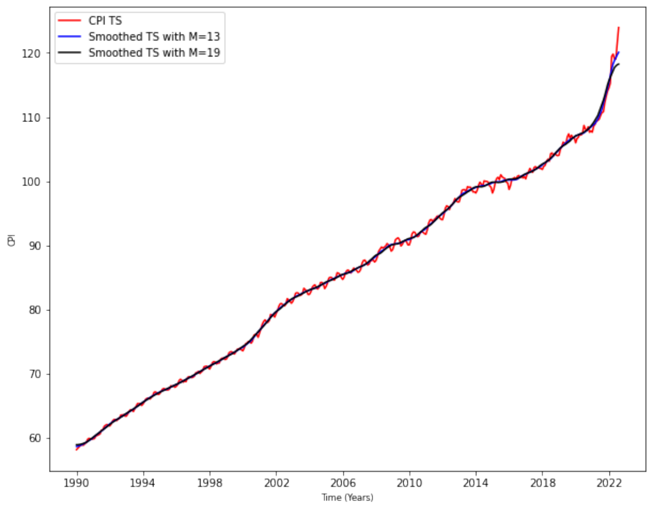

Fig. 3. TS smoothing through Moving Average Filtering

3) LOcally WEighted Scatter plot Smooth (LOWESS): Just like the MA filtering, this approach also uses a local smooth technique to estimate the trend component. However, LOWESS has many steps that, in summary, correspond to calculating weights within a sliding window and updating them by minimizing their error, thus repeatedly improving the smoothing of the TS.
Once again, different M values were tested, 13 and 19, and the resulting smoothed time series are present in Fig. 4.

B. Seasonality Removal

Seasonality relates to recurring patterns within a TS over a certain period of time. These patterns are usually visible in the TS plot, but their periodicity might be harder to estimate by simply looking. By applying the Fourier Transform (FT) to a trendless time series, we can obtain the most common periods of the existing patterns (Fig. 5).

Even though there is a strong component located at 1 cycle/year, which means an annual pattern, the most significant one is at 2 cycles/year, indicating that the most relevant period of seasonality corresponds to 6 months. Now that we established this, we can move on to applying the techniques to remove the seasonality.

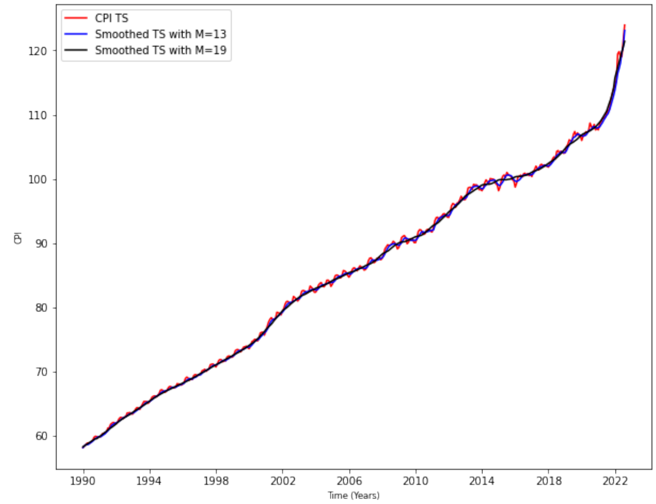

Fig. 4. TS smoothing through LOWESS

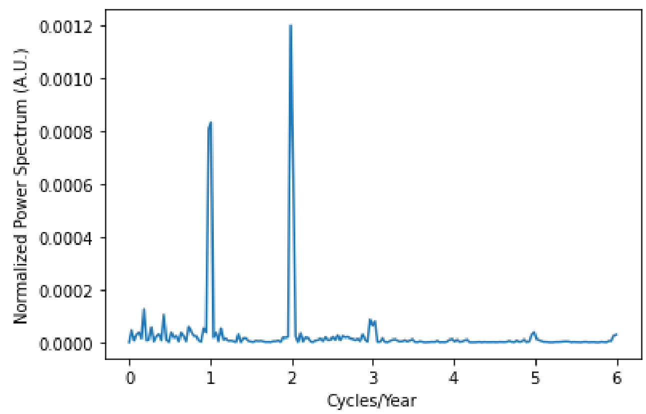

Fig. 5. Frequency Components in the Time Series

1) Filtering: In this technique we apply a filter to remove all but the most relevant frequency component identified (2 cycles/year, visible in Fig. 8). That way, we obtain the seasonality component and are able to remove it from the TS. To do so, we apply a low-pass filter that removes all components over 2.4 cycles/year, followed by removing the resulting seasonality component from the trendless TS, once again assuming a multiplicative model (Fig. 6).

2) Epoch Averaging: This method consists of computing the average cycle within the TS and removing it, repeating the process along the series, obtaining an estimation of the seasonality component. Then, like in the previous method, we can easily remove it, assuming a multiplicative model (Fig. 7).

C. Trend and Seasonality Removal by Differencing

Another approach to achieve stationarity is through differencing, which avoids the need for estimations of trend and seasonality components. That way, we should be able to reach better results. By applying the same principle as the

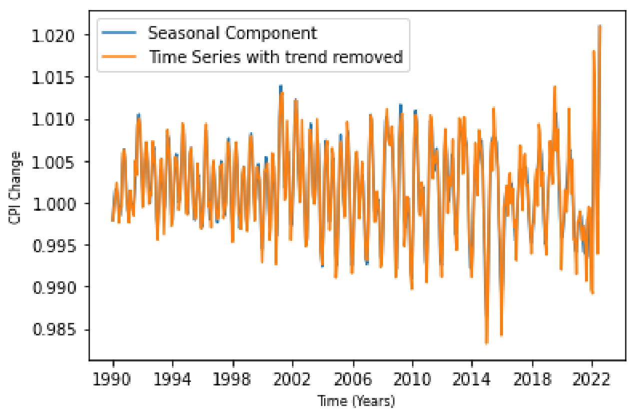

Fig. 6. Seasonality Component calculated through Filtering

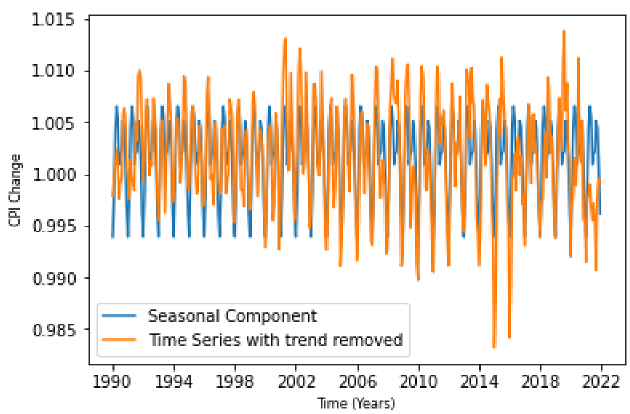

Fig. 7. Seasonality Component calculated through Epoch Averaging

differentiation in mathematics, we are able to remove the trend and seasonality with a first or second order differencing, as we show in Fig. 8.

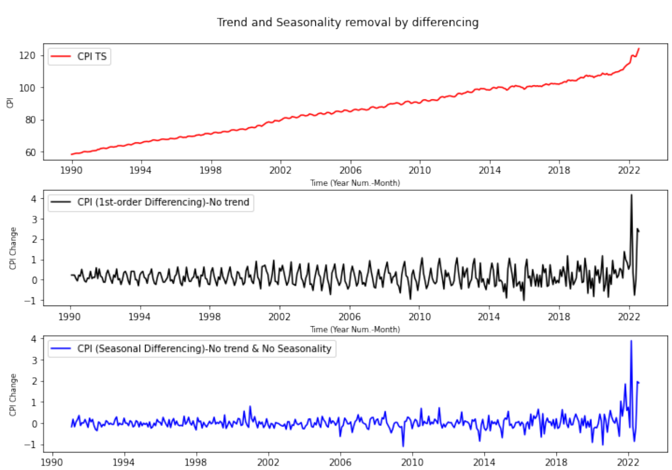

Fig. 8. Trend and Seasonality removal by differencing

D. Stationarity

A TS is considered stationary if the mean and variance are nearly constant overtime, according to the Wide Sense Stationarity (WSS). In other words, if it has no trend or seasonality. This is what we try to achieve in the previous subsections by extracting the erratic component of the TS. Another way to analyze this concept is the through the Autocorrelation, which is, as the name states, the correlation of the TS with itself. If that is low, it means the values do not depend from each other, which hints that there are no such patterns as trend or seasonality. This can be calculated through the Sample Autocorrelation Sequence or Autocorrelation Sequence (ACS):

$$r_{XX}(T) = \frac{\sum_{n=0}^{N-T-1} (x(n) - \bar{x})(x(n+T) - \bar{x})}{\sum_{n=0}^{N-1} (x(n) - \bar{x})^2}$$

By plotting these results over time, we get a correlogram. In Fig. 9 we present the correlogram for the erratic component obtained throught the second order differencing, as described in Subsection II-C.

As we mentioned, we want the ACS to have low values. To analyze this, we set a confidence bounds in order to decide whether or not the values are close enough to zero. By defining a level of significance of 95% (assuming a Gaussian distribution), we consider all values within $\pm\frac{1.96}{\sqrt{N}}$ as zeros. Those bounds are also presentend in the correlogram.

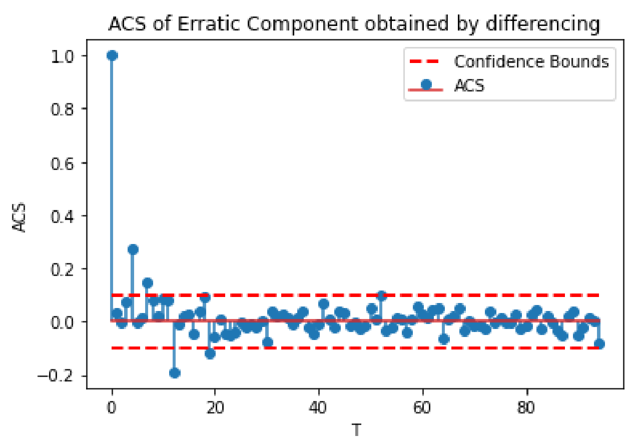

Fig. 9. ACS of Erratic Component obtained by Differencing

III. Results

A. Trend Removal

1) Model Fitting: First of all, we can see in Fig. 2 that the fitting of the linear and quadratic functions is very similar (so much that we can't even see the linear fitting plot), just as the resulting TS after removing the trend. We also tested a higher order polynomial fit to try to better adjust to the TS. However, in all three cases, we can still observe multiple local trends, which indicates this is not a great solution.

2) Moving Average Filtering: As seen in Fig. 3, the two smoothed TS's are very similar. We decided to choose the one generated by M = 19 to reduce a possible overfitting. Then, after removing that trend, we obtain the TS presented in Fig. 10.

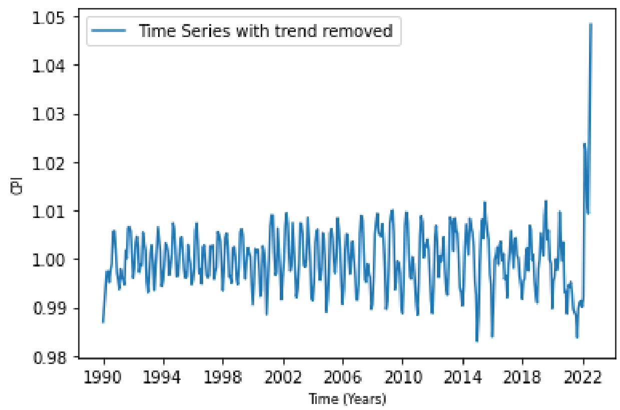

Fig. 10. TS after trend removal through MA Filtering

While the global trend has been successfully removed, we can still observe some local trends, although reduced.

3) LOcally WEighted Scatter plot Smooth (LOWESS): By looking at Fig. 3 we see that the smoothed TS for M = 13 slightly overfits the original one. Therefore, we chose the one with M = 19. The resulting TS after the trend removal is presented in Fig. 11.

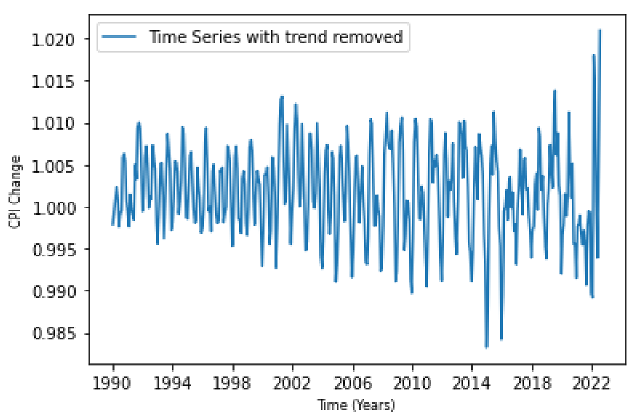

Fig. 11. TS after trend removal through LOWESS

Just as with the MA filtering, the global trend is removed but some local trends can still be seen.

From the three approaches presented, there is a clear difference between the Model Fitting and the others, that perform much better. Between MA Filtering and LOWESS though, the results are very similar, yet we believe the latter has a slight advantage. However, we believe there are better approaches than these, namely the differencing method.

B. Seasonality Removal

1) Filtering: Even though we established that the main frequency component was 2 cycles/year, we also tested seasonality analysis with 3 cycles/year since the resulting TS obtained still showed some visible patterns. The estimated seasonality component fitted the TS better and had a smaller error (although very little difference).

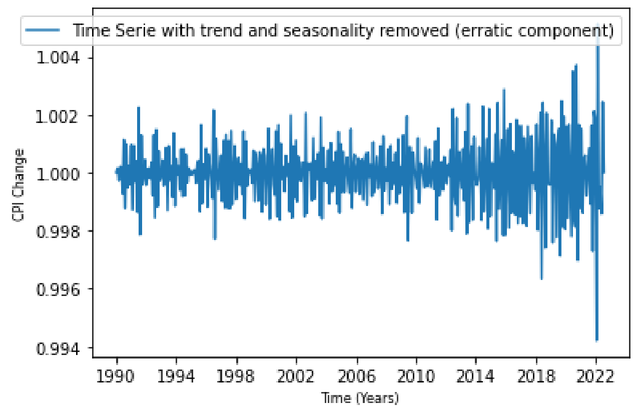

Fig. 12. TS after trend and seasonality removal through LOWESS and Filtering

We are satisfied with these results since we cannot identify any patterns.

2) Epoch Averaging: As we saw in Fig. 7, the estimated seasonality component does not seem to fit the TS very well. Because of that, the poor results obtained are not very surprising: many patterns can be identified.

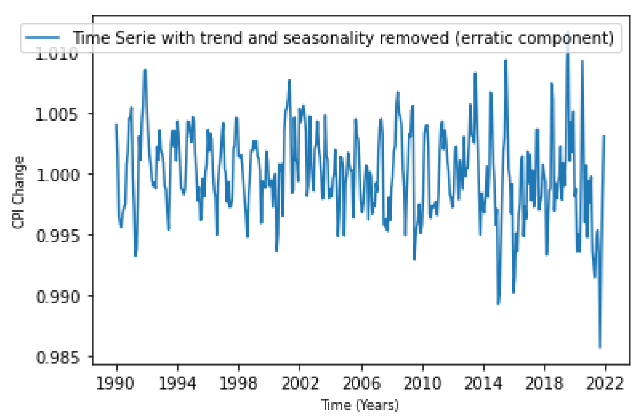

Fig. 13. TS after trend seasonality removal through LOWESS and Epoch Averaging

C. Trend and Seasonality Removal by Differencing

As we can see in Fig. 8 we were able to successfully remove the trend by using the first order differencing and the seasonality through the second order differencing, only remaining the erratic component, which appears to be very noisy and random, as expected. As we previously explained, these good results are obtained from avoiding the calculation of estimations to remove the other components. Therefore, we believe this is the best strategy to reach stationarity, from the ones we tested. However, this analysis is done more carefully in the following subsection.

D. Stationarity

In Fig. 9 we see the correlogram for the erratic component obtained throught the second order differencing. This was the only resulting TS analyzed in terms of stationarity since it presented the best results visually.

Looking at the ACS values, we find that most of them fall within the confidence bounds, except for a few points. Those exceptions might be related to a small seasonality not removed by the second order differencing. Nevertheless, we believe these are satisfactory results.

IV. Discussion

To be written in the future.