Consumer Price Index Analysis and Forecasting

Pedro sá
University of Coimbra
Coimbra, Portugal
pedronuno@student.dei.uc.pt

Alexandre Santos
University of Coimbra
Coimbra, Portugal
alexandresantos@student.dei.uc.pt

Abstract—The Consumer Price Index (CPI) is a macroeconomic indicator that measures price changes for goods and services purchased by households to meet a set of desires and requirements, and is commonly leveraged by governments for estimating inflation and policy-making. A proper modelling of the CPI can be key for better decision-making on the economic state of a country. In this paper, we comprehensively analyze and forecast Denmark's CPI.

Index Terms—time series, forecasting

I. INTRODUCTION

The Consumer Price Index (CPI) is a macroeconomic metric of the price changes for goods and services acquired by households to satisfy a set of wants and needs. This index is important for decision-making processes in governments and institutions, e.g., for providing precise estimates of inflation, and thus poses a handy reflection of the general economic state and well-being of a country.

An accurate modelling of a country's CPI is key to understanding out it behaves now, and how it will behave in the future. Moreover, it's unquestionably useful for predicting periods of inflation, allowing governments to better handle market dynamics before they happen.

In this paper, we present a case study on the country of Denmark, comprising an analysis and forecasting model of the CPI.

A. Related Work

B. Data

In order to perform such study, a dataset consisting of CPI (and several other indicators) records was made available. From the 191 available countries, we selected Denmark as a subject of study and therefore extracted a time series of monthly CPI (all items) values from 1990 up until 2022.

The time series was split into a training and test set, for which consider only 12 samples (the last year) for testing. Figure 1 illustrates the evolution of the CPI over a span of 22 years for the country of Denmark.

II. METHODS

The following section describes in detail the methods used for characterization, modelling and forecasting of the time series.
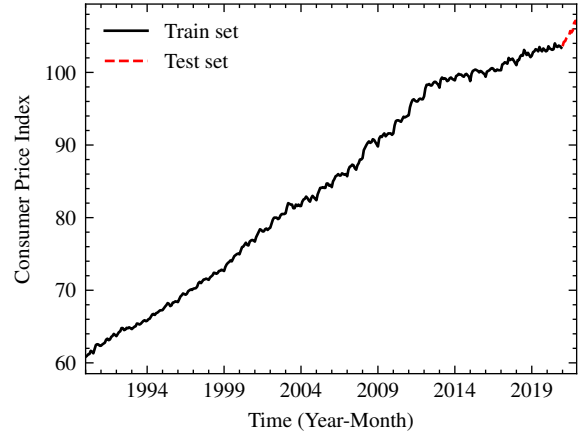

Fig. 1. Denmark CPI (All Items)

A. Characterization

Preliminary observations suggest that the time series better fits a multiplicative model of the form:

$$ts(n) = tr(n) \cdot sn(n) \cdot \epsilon(n)$$

given the seeming independence of the seasonality and erratic component from the "apparent" trend.

1) Trend: The trend component was described by three different techniques. **Polynomial fitting** is the process of fitting a polynomial curve of order n

$$tr(n) = \beta_0 + \beta_1 n + \beta_2 n^2 \cdots + \beta_k n^k$$

for which the coefficients $\{\beta_i\}_{i=1}^k$ are estimated by minimizing a l_2 cost function

$$\min \sum \Big(x(n) - tr(n) \Big)^2$$

with respect to the residuals. Linear and Quadratic fitting were used in this paper.

Moving-Average is a local smoothing technique that averages the values on a *moving* time window M along the timeseries,

$$\hat{tr}(n) = \frac{1}{\sum w_k} \sum_{n-\frac{M-1}{2}}^{n+\frac{M-1}{2}} \omega_k x(n+k)$$

thus removing some randomness from the data. The moving average implements data augmentation to account for the non

existent samples, at the beginning and at the end of timeseries, needed to calculate the average.

LOcally WEighted Scatterplot Smoothing (LOWESS) is a non-parametric approach that performs a local weighted regression over the M closest points to each data point (x_i, y_i). First, the regression is fit by a weighted l_2 cost function, giving more weight to nearby data points. In order to decrease the effect of outliers, further regressions can be performed. In the experimental setup we chose to perform only 1 additional regression.

N-order differencing is a technique for *implicitly* turning non-stationary time series into stationary, by calculating the changes between consecutive observations. The first difference is given by

$$t\acute{s}(n) = ts(n) - ts(n-1)$$

and allows to remove the trend-cycle. Seasonal differencing is of the form

$$t\acute{s}(n) = ts(n) - ts(n-m)$$

in which m is the frequency of a seasonal pattern. The latter is used to remove any trace of seasonality and attain stationarity if the first-order difference was not enough.

2) Seasonality: Besides the above mentioned **n-order differencing**, there are additional techniques to assess seasonality. **Spectral Analysis** is a technique for finding underlying periodicity in stochastic processes; by transforming a time series from the time domain to the frequency domain, we can characterize it on the basis of its distinct frequencies – seasonal patterns. The frequency spectrum is calculated by the **Discrete Fourier Transform**, a technique that decomposes a complex signal in more simple ones, which will yielded the different frequencies – seasonal patterns – and corresponding magnitudes.

Filtering builds on top of spectral analysis, because upon isolating the cycles, these can be removed by filtering the frequencies (or seasonal patterns) of smaller magnitude as to provide a more accurate estimate of the cyclic behaviour. The filter used in this paper is the Butterworth.

3) Stationarity: Along with **n-order differencing**, stationarity can also be assessed via other methods. **Autocorrelation analysis** consists of computing the correlation r_{XX} between observations at different time lags, for which $|r_{XX}| > 0$ indicates some correlation and $r_{XX} = 0$ indicates an absence of correlation. By establishing confidence bounds, considering a significance level of 95%, we can say that the autocorrelation is bounded by

$$-\frac{1.96}{\sqrt{N}} \le r_{XX} \le \frac{1.96}{\sqrt{N}}$$

where values that fall outside this region are *significantly* different from zero and indicate a substantial correlation. For a non-stationary time series, the following is verified

$$\lim_{T \to \infty} r_{XX} = 0$$

TABLE I
TREND ASSESSMENT METHODS

Method	Parameters	MSE
Linear	–	2.47e − 04
Quadratic	–	2.43e − 04
Moving Average	$M = 13$	1.29e − 05
LOWESS	$M = 13$	1.09e − 05

whereas if the autocorrelation quickly drops to zero, it generally means that the time series is stationary. The **Augmented Dickey-Fuller test** is a statistical test that tests the null hypotheses H_0 that a unit root is present in a time series. A time series is considered to be non stationary if a unit root is present and stationary otherwise.

B. Modelling

For the modelling, considering the time series is indeed non-stationary and exhibits cyclic behaviour, we use a SARIMA model. A SARIMA model extends the ARIMA model with seasonal terms:

$$\text{SARIMA}(p, d, q)\text{X}(P, D, Q)_S$$

where d and D stand for the order of first and seasonal differencing, respectively; p and P represent the Autoregressive (AR) model orders for the non-seasonal and seasonal part of the model, respectively; q and Q stand for the Moving Average (MA) model orders for the non-seasonal and seasonal part of the model, respectively, and S represents the dominant cyclic behaviour.

C. Forecasting

For the forecasting task, we consider both the previous SARIMA model of the timeseries, and also the Holt-Winters Exponential Smoothing (or Triple Exponential Smoothing).
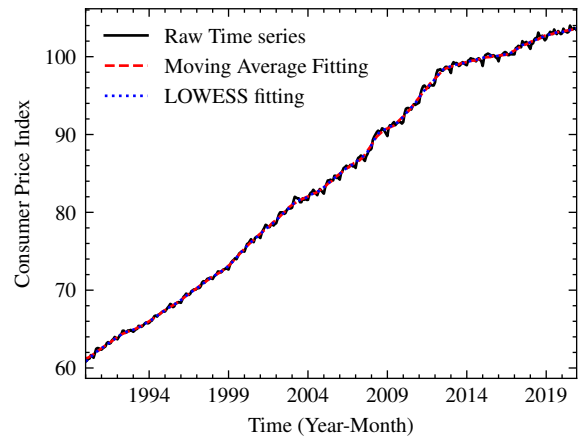
III. RESULTS

A. Characterization

Fig. 2. Trend fitting

The results suggest that the time series's trend is better described by the LOWESS approach as seen in Table 2 and

Subject	p-value	ADF Statistic	1% Treshold
Raw timeseries	0.645	-1.266	-3.449
First difference	0.042	-2.928	-3.449
Seasonal difference	≈ 0	-7.937	-3.449

Parameter	Range	Best
SARIMA		
p	$[0, 2[$	1
q	$[0, 2[$	1
P	$[0, 4[$	0
Q	$[0, 3[$	1
Best AIC	≈ -191.7	
Best $\mathcal{L}$	≈ 99.8	
Q-statistic	≈ 0.09	
Holt-Winters		
α	$[0.1, 0.9[$	0.8
δ	$-$	0.03
γ	$[0.1, 0.9[$	≈ 0.3
Best AIC	≈ -1145	

Figure 2. The linear and quadratic trends, due to their low-order nature and usage of the global structure of the data, are not sensitive to oscillations and thus are only able to give a rough approximation. The Moving Average and LOWESS fittings, on the other hand, rely on the local structure of the data and are globally non-linear, allowing them to tend to these oscillations. LOWESS yields a better approximation as it performs multiple local (weighted) regression instead of averages, thus being more sensitive to outliers.

Regarding seasonality, a careful glance suggests a clear pattern every 6 months and, less noticeably, every year. Through spectral analysis, we we're able to identify 1 and 2 cycles per year, as seen in Fig 3, and thus we filtered the other non relevant frequencies. The estimated seasonality is also illustrated in Fig 3, from which we can observe an adequate estimation of the cyclic behaviour, given that it models both cycles fairly well. Additionally, the erratic component resembles some stationarity, although some patterns still remain visible; this most likely stems from this method relying in estimations and decomposition models.

The differencing approach on the other hand, yielded more clarifying results. This may be due to differencing not resorting to the estimation of each individual component, and instead computing the differences between data points and removing the changes in level. As seen in Figure 4, the first difference was not enough to achieve stationarity given some clear patterns and a slight upwards slope, whereas the seasonal difference time series has constant variance and is arguably horizontal. Observing the correlogram in Fig 5, we can argue for non-stationarity of the first difference, considering i) most values fall outside of the confidence bounds ii) a strong decreasing autocorrelation for lag values multiple of 6 and 12, which happens to be the most relevant seasonal patterns in the timeseries. On the other hand, the seasonal difference is clearly stationary, given that i) almost all values tend to be within the confidence bounds, apart from a few exceptions[1] ii) the autocorrelation instantly drops to zero.

Table II presents the results of the ADF unit root test. As expected, H_0 is accepted for the raw time series, as p-value > 0.05, whereas it's actually rejected for the first and seasonal differences, considering the p-value is inferior to 0.05 and that the test statistic is below the 1% confidence level for the *seasonal difference*; for the *first difference*, the p-value is acceptable but the statistic is actually above between the 1% and 5% confidence level. While this poses a contradiction for the first difference, one should take into account that the ADF test actually tests if there is or isn't a need for further

[1]Probably due to computational errors given the small sample size N

differencing, thus suggesting that the first difference doens't require any more differencing, which can be construed as stationarity.

B. Model Evaluation

These kind of assesments allow us to derive the parameters for time series modelling. Because stationary was only attained with a first difference and a seasonal difference, we can set $(d = 1; D = 1; S = 12)$. Moreover, from the ACF and PACF (see Appendix A) we can guess the AR and MA parameters of the model. Instead, we opted by a grid-search approach for finding the parameters, considering the ACF and PACF are often times deceiving, as the parameters we derived from a visual analysis yielded a better AIC score (≈ -197.1). Table III summarizes the results attained.
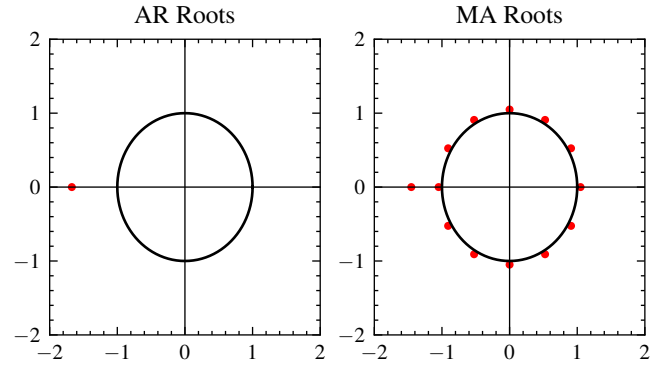

Fig. 6. Characteristic roots of the AR and MA processes

Figure 6 illustrates the characteristic roots for the model fitted to the time series. Both the AR and MA polynomial roots happen to lie outside the unit circle, which is mandatory for preserving the stationarity and invertibility properties of the model. However, the MA roots lie too close to the unit circle: these can become numerically unstable and eventually lead to poor forecasts, as we later discuss.
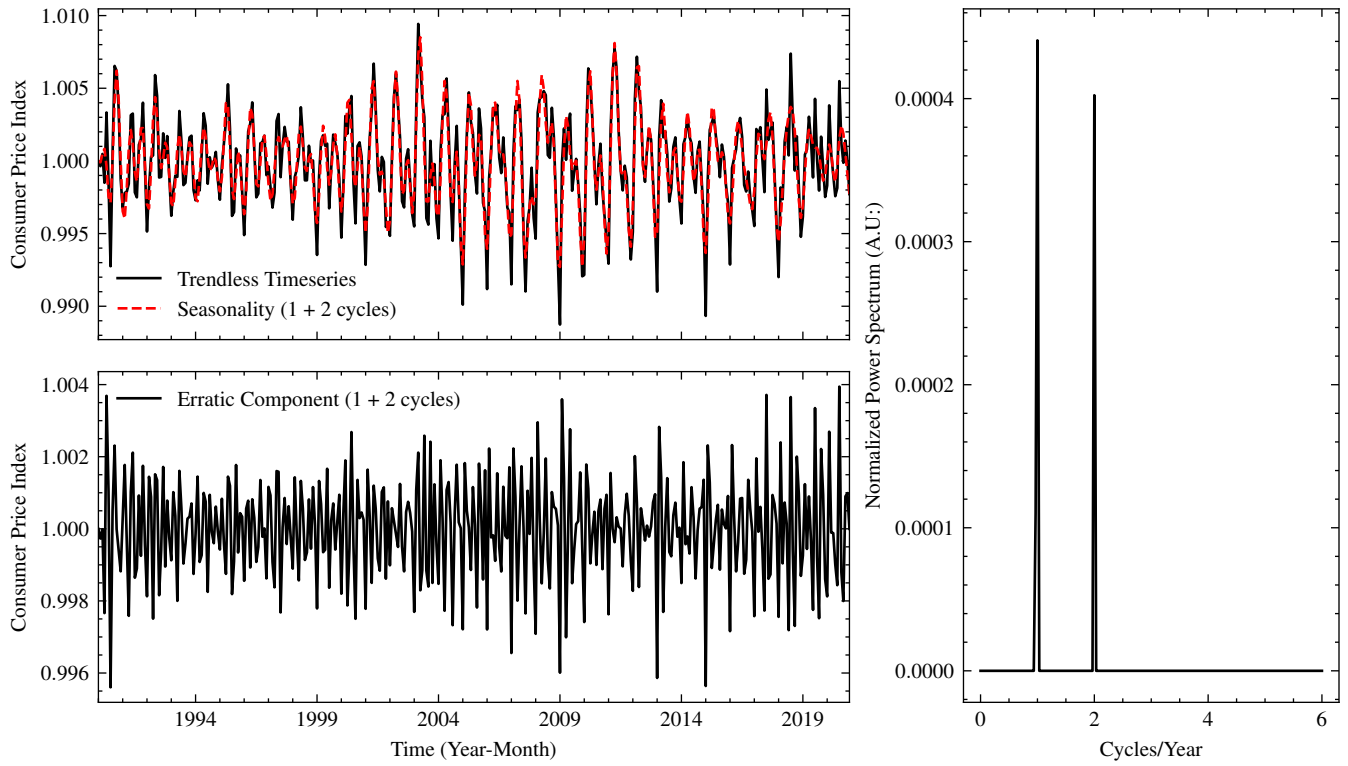
Fig. 3. The frequency spectrum (right), the trendless timeseries and the seasonality fitting (top left), the estimate erratic component (bottom left).
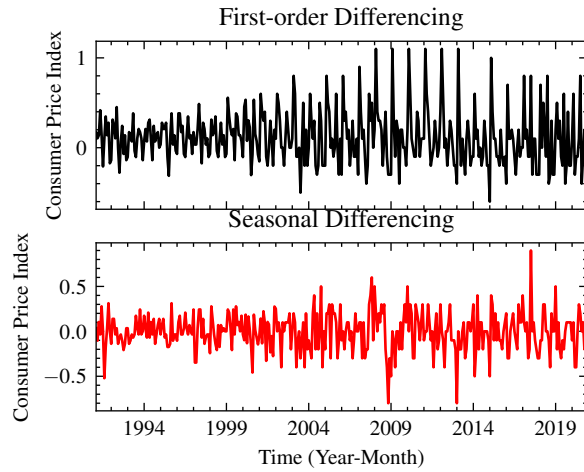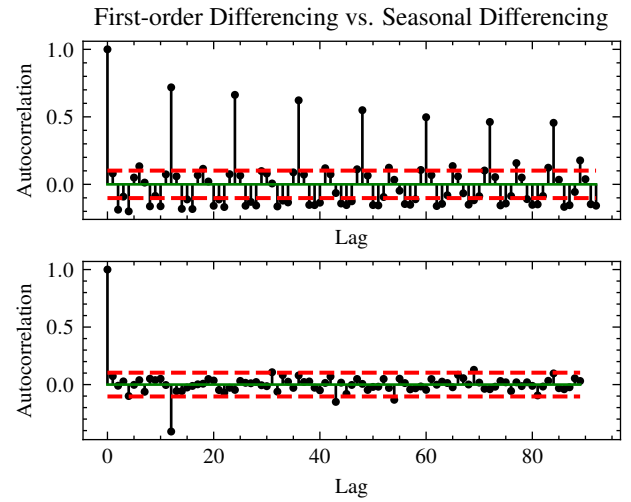

Fig. 4. Stationarity by differencing

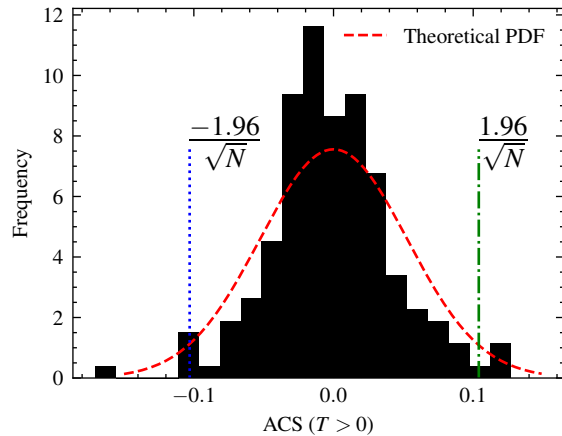Fig. 5. Correlogram of first-order differencing (top) and seasonal differencing (bottom)

Figure 7 illustrates the residuals' ACF values distribution, which arguably follows a normal distribution. Moreover, it seems that the number of available samples has relaxed the confidence bounds to a broader range. It so happens to be that the available data for training consists of only $N = 372$ samples, which, and according to the Law of Large Numbers, doesn't provide enough statistical significance for the model to approximate fittings close the expected value. Such is also

Fig. 7. Distribution of the residuals' ACF

TABLE IV
UNIVARIATE FORECASTING RESULTS

Model	RMSE ↓
SARIMA	≈ 1.508
Holt-Winters	≈ 1.291

TABLE V
MULTIVARIATE FORECASTING RESULTS

Model	RMSE ↓
MLP	≈ 0.111

evident by the close, but not quite right, approximation to the theoretical probability density function.

Regarding the Holt-Winters Exponential Smoothing approach, we found the best parameters to be the ones detailed on Table III.
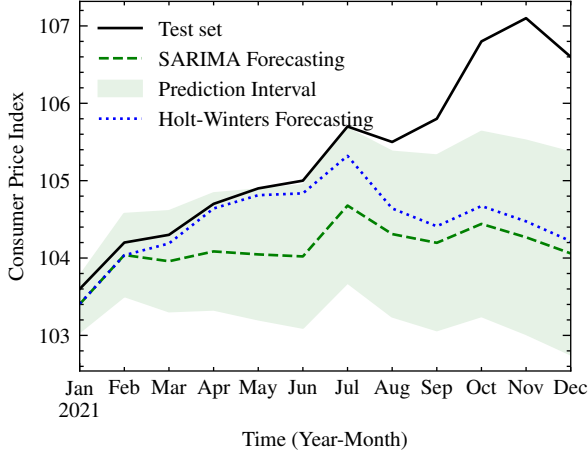
C. Forecasting

Fig. 8. Forecasting on a 12-month period

For forecasting purposes, we found the Holt-Winters method to actually be superior on the test set, see Table IV. As illustrated by Figure 8 they both seem to capture seasonality adequately, but Holt-Winters seems more sensitive to the underlying trend. This is most likely due to how Holt-Winters expresses a process, i.e. as a (in this case, additive) composition of explicit individual components.

D. Multivariate Forecasting

Moreover, we compare the time series of Denmark to that of Norway and Finland to uncover (or not) any relationship between them. We split the data into a training, validation and testing set, saving just 4 years for validation and 1 year (the last) for testing.

By the cross-correlation function between the different timeseries we find that Denmark and Finland are the most co-related ones and for the rest the little existent correlation is non-significative for larger lag values ($T > 0$). Moreover, there seems to be a spike of correlation for lag of periods of 12, which suggests that the timeseries exhibit cyclic behaviour with the same frequency and period.
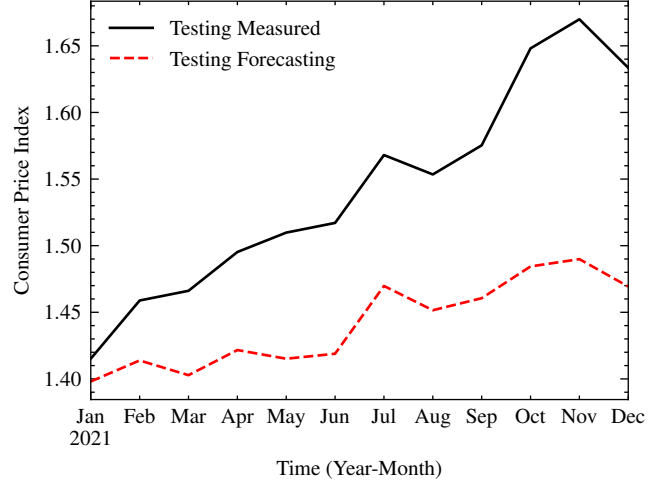

Fig. 10. Multivariate forecasting on a 12-month period

In this approach, we used an MLP to regress the values of the CPI based on features extracted from time-lagged versions of the Norway and Finland timeseries. As portrayed in Table V, the MLP actually attained an arguably low RMSE, considering time-delays of $[0, 1, 12]$ for both Norway and Finland and $[1, 2, 12]$ for itself, Denmark. Figure 10 illustrates the fitting by the model. Note that the scale of the CPI doesn't match that of the univariate forecasting as the MLP requires the data to be standardized. As such, it would be innapropriate to compare it to the univariate forecasting, besides the forecasting closely resembling the SARIMA model: the SARIMA error for the highest peak of the test set is $\epsilon = \approx 3$, while for the MLP is just $\epsilon = \approx 0.3$.

IV. DISCUSSION

In this study we present a thorough analysis of the CPI for the country of Denmark. We began by providing a characterization of timeseries, for which we were able to uncover trend and cyclic behaviour. We then provided a comparison of two well-known models for timeseries, and found Holt-Winters to the best at forecasting. Additionally, we also provided a multivariate approach to forecasting with a MLP, for which we considered two additional countries to uncover realistic relationships and predictive capabilities.

APPENDIX

The following provides an analysis of the ACF and PACF of the timeseries for derivation of parameters for the SARIMA model. As previously mentioned, the raw timeseries presents a trend and clear seasonality. After characterization (see III-A), we found that a first difference followed by a seasonal difference can achieve stationarity, so the parameters ($d = 1; D =$
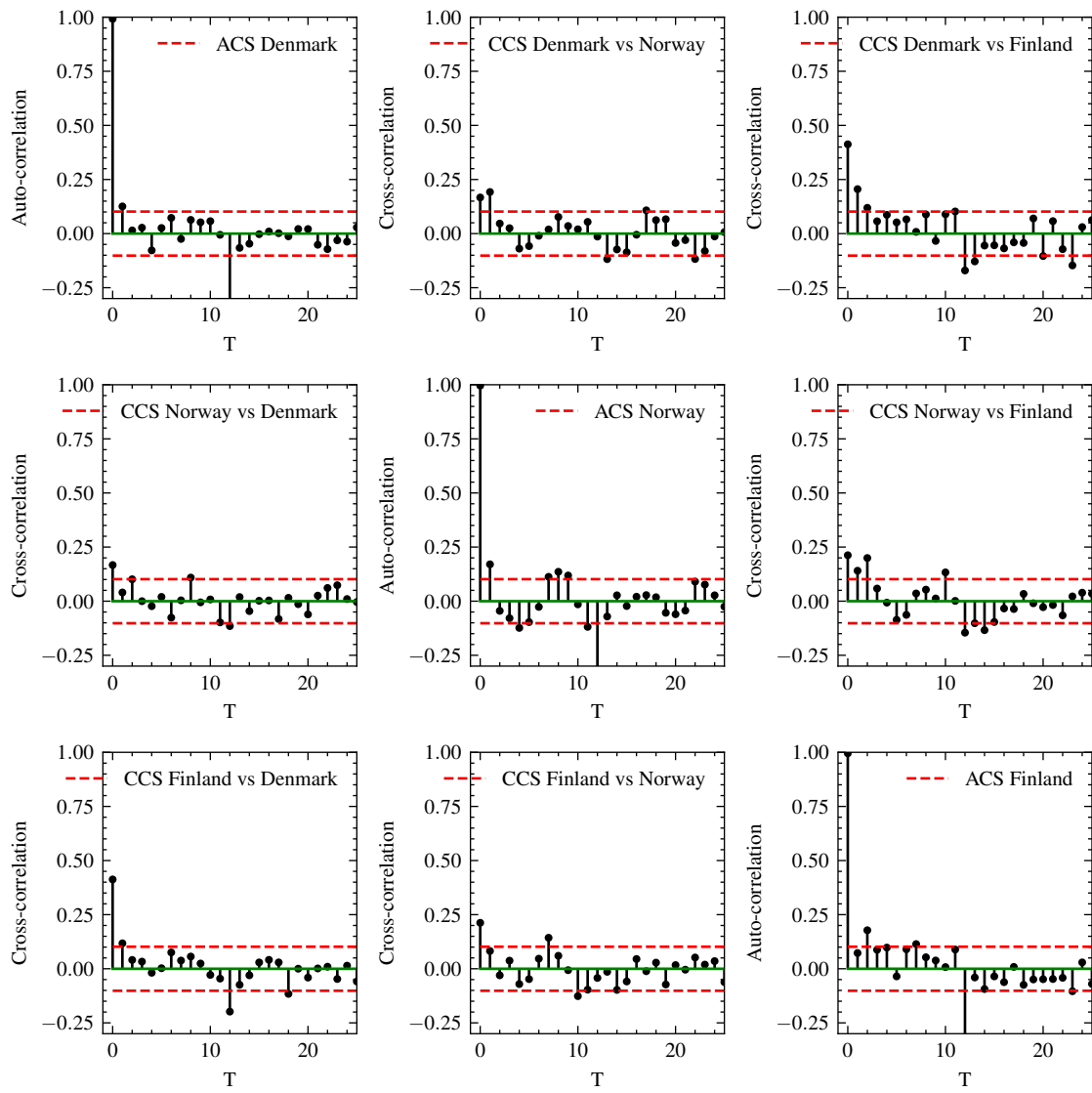
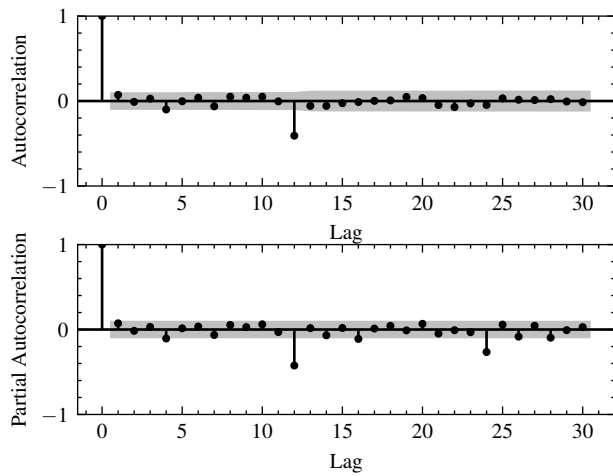Fig. 9. Cross-correlation sequence for timeseries of different countries

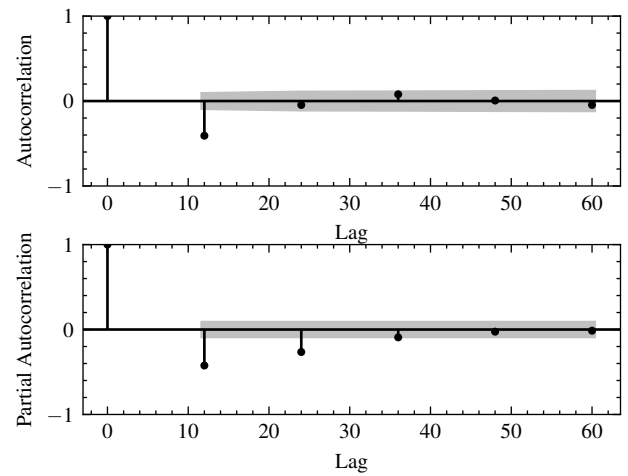Fig. 11. ACF and PACF of the stationary timeseries

Fig. 12. Sample ACF and PACF of the stationary timeseries

$1; S = 12$) are fixed. Figure 11 shows that the auto-correlation quickly drops after $T = 0$, so we fix $(p = 0; q = 0)$. Figure 12 pictures the sampled auto-correlation and partial auto-correlation and suggests that the former drops after the first seasonal period, whilst the latter only reaches values close to 0 2 cycles later. Therefore, we fix parameters $(Q = 1; P = 1)$, but we could be a bit more flexible and accept $P = 2$ or 3.

This parameter derivation fixes our SARIMA model at

$$\text{SARIMA}(0, 1, 0) X(1, 1, 1, 12)$$

which yielded an AIC of ≈ -189.3 which is actually close to the optimal value we found (≈ -191.7). This suggests that the ACF and PACF plots leave a considerable room for interpretation.