Real Gross Domestic Product for Italy

Massimiliano Pastorino and Caterina Penna

The purpose of this study is to analyze the time series related to Italian GDP. The time series was retrieved from the website of the Federal Reserve Bank of St. Louis.
Here is the url:

https://fred.stlouisfed.org/series/NGDPRNSAXDCITQ

The data considered range from the first quarter of the year 1996 to the first quarter of the year 2022. We will first analyze the main features of the time series by extracting the three components (trend-cycle, seasonality, and residual) of a time series by applying classical decomposition and STL decomposition. Next, the goal is to identify the process that best represents the series.

1 Characteristics of Time Series

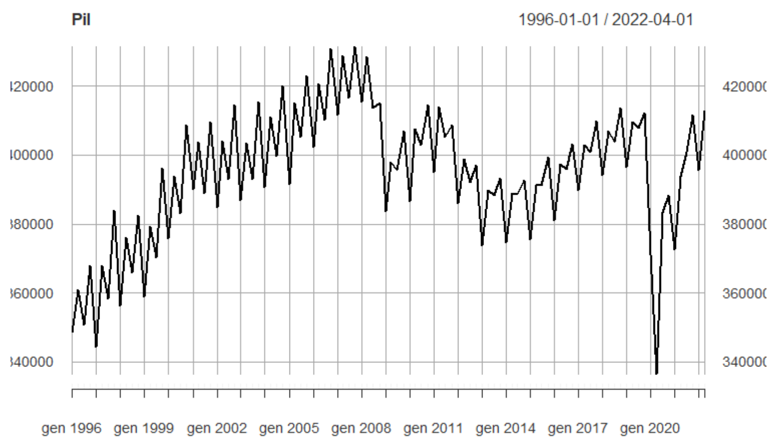

Figure 1: **Time plots**

From an initial visualization of the data we can infer that the time series has a strong seasonality, we have peaks that repeat periodically at Q2 and Q4 each year, while in Q1 and Q3 we can see a decline.

In addition, the series is characterized by an increasing trend at times with some sudden reversals in the medium to long term, in fact from the time-plot it is evident that up to 2008 we have a sharply increasing trend, thereafter we have a fluctuating and uneven trend as we can see the alternation of the increasing and decreasing trend.

The actual data may contain missing values and outliers. Outliers are observations that deviate greatly from other observations in the time series.

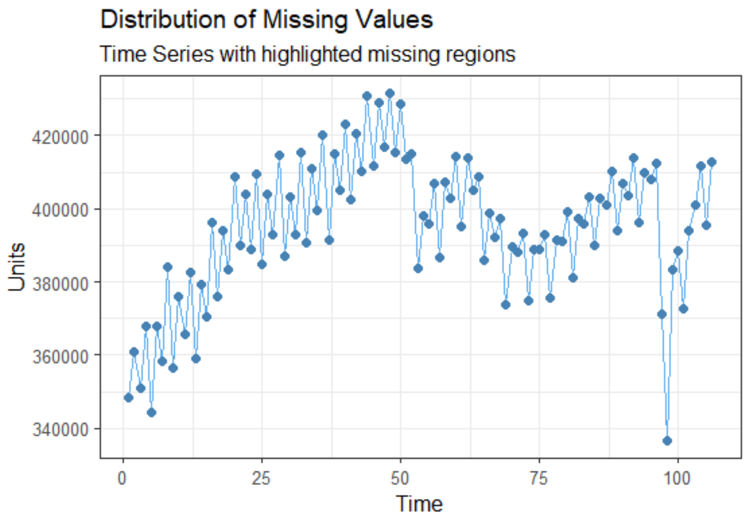

Figure 2: **Distribution of Missing Values**

Figure 2 shows the distribution of missing values; from the graph it can be seen that the time series has no missing values.

Table 1: **Possible outliers**

DATE	PIL
2019-07-01	407937
2019-10-01	412337
2020-04-01	336460

From Table 1 we can see the presence of outliers classified as potential outliers. Compared to most values assumed by the time series, the values identified as potential outliers are simply unusual values.

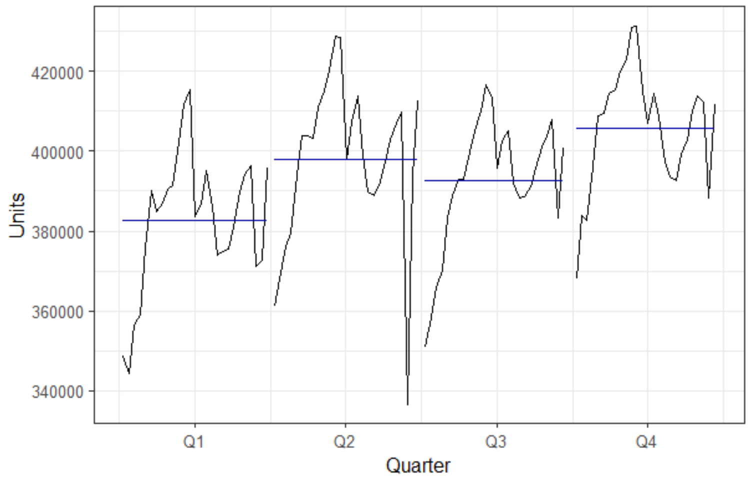

Figure 3: **Seasonal subseries plots**

This graph is useful for identifying whether there is a trend in seasonality in the graph. The horizontal lines indicate the averages for each quarter, so the first line indicates the average value referring to the 1st quarter for all years, the second line indicates the average value referring to the 2nd quarter for all years, and so on.

We can confirm what was said initially and that is that all the years in Q1 and Q3 show a decline, in fact they have a lower average GDP than Q2 and Q4 which as said show an increase; also we can see that Q4 has the highest average GDP than all the other quarters. Given that the average varies and does not remain constant we can say that there is seasonality in the data.

1.1 Trend and seasonality in ACF/PACF plots

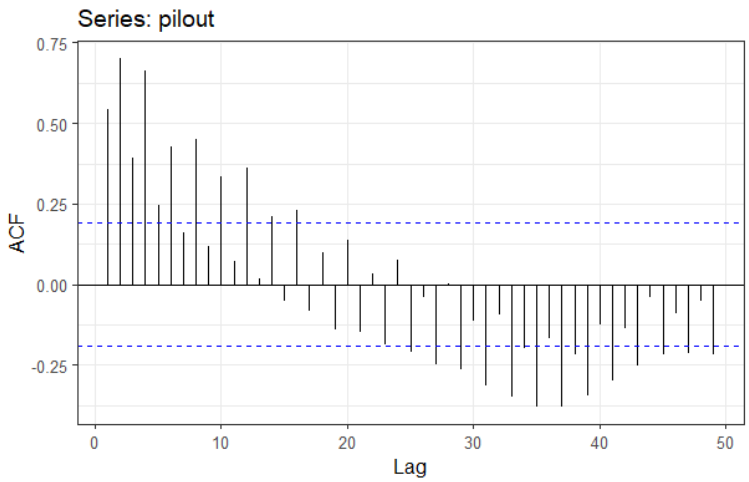

Figure 4: **ACF plots**

Autocorrelation measures the extent of a linear relationship between lagged values in a time series. In the ACF plot we have 95% confidence bands that are useful to assess whether the autocorrelations are significantly different from 0. We observe the highest autocorrelation value at lag 2, while the value of the autocorrelation at lag 0 is always 1, we can also see that the series is not White Noise since most of the autocorrelations come out of the confindence band.

In fact in a White Noise series we expect the autocorrelations to be close to 0, however, taking into account the component of error due to random variance, referring to the 95% confidence interval, we expect that the time series is White Noise if the 95 % of the values are within the bands, this does not happen and therefore the series is not White Noise.

The trend and seasonality are also evident in this graph, in fact the decrease of values in the ACF as the lag increases is due to the trend, while the wave shape is due to seasonality; there are also peaks present for the lag 2,4,6,8,... and so on so we have a seasonality of length 2. The series is not stationary as the ACF decreases slowly.

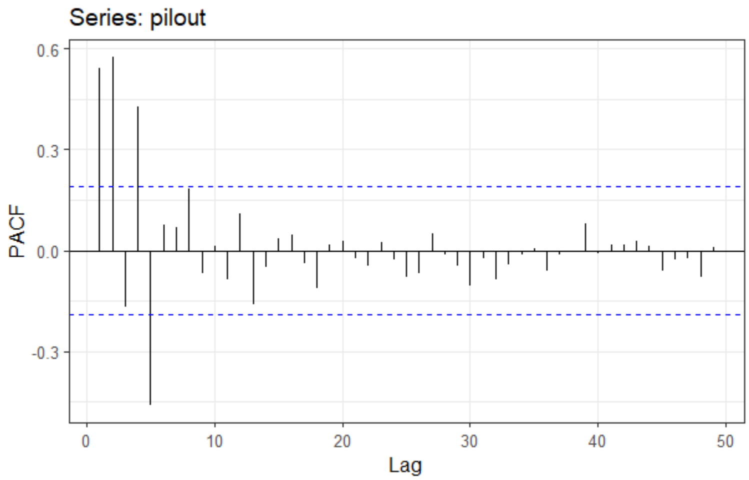

Figure 5: **PACF plots**

The partial autocorrelation plot, shows the correlation after removing the effect of any correlations due to the shortest lag terms, in fact the PACF graph has significant peaks only for the earliest lags, which means that the higher-order autocorrelations are indeed explained by the autocorrelations of the smallest lags.

2 Time series decomposition

By focusing on the components of the series, we want to capture the trend component through the use of a simple centered moving average. The optimal average is one of order 4.

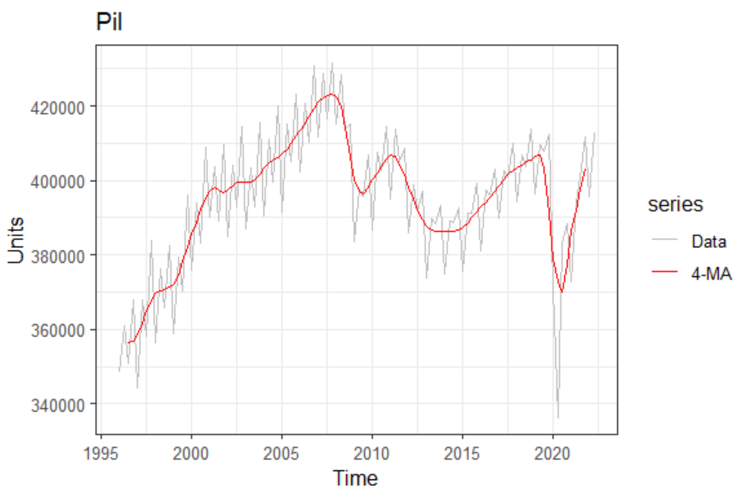

Figure 6: **Moving averages**

The moving average with this order is able to reproduce the underlying trend of the series and is also able to capture the level changes in the years 2008, 2012 and the collapse which occurred in 2020.

- The order of the moving average determines the degree of smoothness of the series. The greater the order, the smoother the series will be. Conversely, the smaller the order, the more closely the series will follow the data and be more uneven.
- Moving averages are usually of odd order, to be symmetrical. To make an even-order moving average symmetrical, a moving average can be applied to the moving average. For example, you can take a moving average of order four and then apply another moving average of order two to the results to obtain a symmetrical series.
- In this case, the center(in R) parameter was used to center the moving average, so the approach described above described represents only an alternative.

2.1 Classical decomposition

Another technique used in this analysis is decomposition, which is a process that aims to decompose the series into its components to understand which of these most determine the evolution of the phenomenon over time.

- Trend: determines the underlying trend or long-term trend period of the series.
- Seasonality: determines the fluctuations that are repeated with regularity over time.
- Residuals: anything that cannot be captured by the model.

In the decomposition of the series, the cyclic and trend components are considered together, which is why it is referred to as the cyclic trend component. The first decomposition technique used is the classical technique, which can be additive or multiplicative. These techniques are simple to apply but have limitations, i.e., they do not provide a complete estimate of the trend component and assume that the seasonal component is fixed, obviously this can be a constraint for series that have a variable seasonal component.

2.1.1 Additive decomposition

In additive decomposition, the original time series is assumed to be defined as the sum of the three components (trend, seasonality and residuals). It consists of the following steps:

- Step 1: Estimate the cyclical trend component with the use of the centered moving average.
- Step 2: Calculate the "detrended" series and subtract the cyclical trend component from the original series.
- Step 3: Estimate the seasonal component using the series "dentrendized" series. The seasonal component for each season is estimated as the average of the detrended values for that season. In this case, with quarterly data, the component seasonal for a quarter is the average of all values detrended for that quarter.
- Step 4: Estimate the residual component. It can be obtained by subtracting from the original series the component of estimated trend/cyclical and the seasonal component.

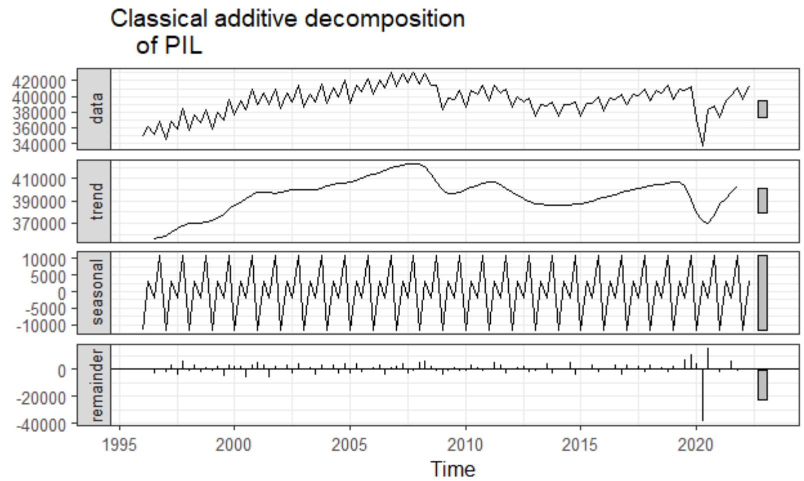

Figure 7: **A classical additive decomposition of PIL**

The graphs represent the observed time series, the trend component, the seasonal component, and finally the residual component, respectively.

- The estimated seasonal component has periodic and regular fluctuations with constant amplitude; that is, seasonality is constant from year to year. This reflects the assumption on seasonality underlying the classical decomposition, which precisely assumes that the seasonal component is constant over time(from year to year).
- The gray bar shows the relative magnitude of the effects. The larger it is, the smaller the variation in the component will be relative to the observed series.
- The residual component has zero mean as the fluctuations fluctuate around zero. We have more or less constant variations.

By comparing these two components, we are able to see where the model fails to fully capture the structure of the data. As we can see from Figure 8, the cyclical trend component estimated by the additive method captures sudden changes(upward or downward) quite well. However, the component seems to be too smooth, so it is too homogeneous when there are sudden upward or downward variations in the data.

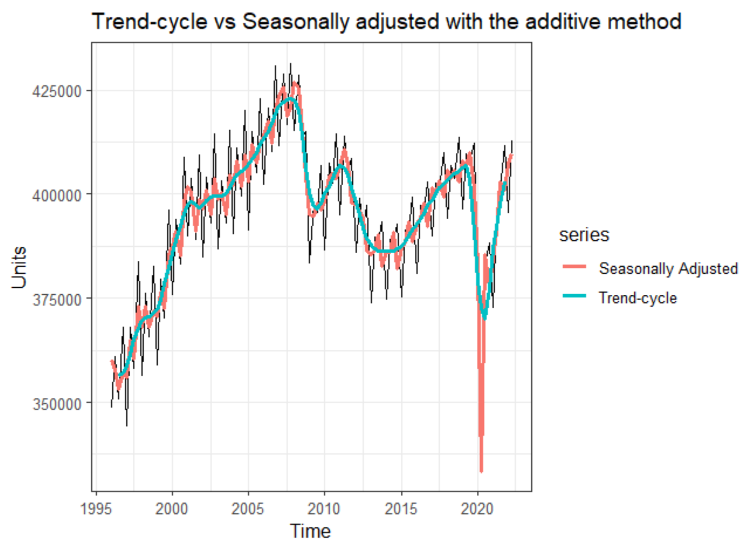

Figure 8: **trend-cycle vs seasonally adjusted additive model**

2.2 STL decomposition (Seasonal and Trend decomposition using Loess)

STL decomposition is a versatile method of decomposing the series; it has several advantages over classical decomposition: it handles any kind of seasonality, the seasonal component can change over time and the rate of change can be controlled, the "smoothness" of the trend-cycle can be controlled, and finally it can be robust to outliers, so that unusual and occasional observations do not affect the estimates of the trend-cycle and seasonal components (they will still affect the residual component).

On the other hand, STL has some disadvantages, including: it does not automatically handle the variation in the day. The parameters t.window and s.window control how rapidly the trend-cycle and seasonal components can change. Smaller values allow for more rapid changes. t.window is the number of consecutive observations to be used when estimating the trend-cycle while s.window is the number of consecutive years to be used in estimating each value in the seasonal component. s.window must specify as there is no default. With s.window = 'periodic' we assume that the seasonal component is no longer variable but constant over time.

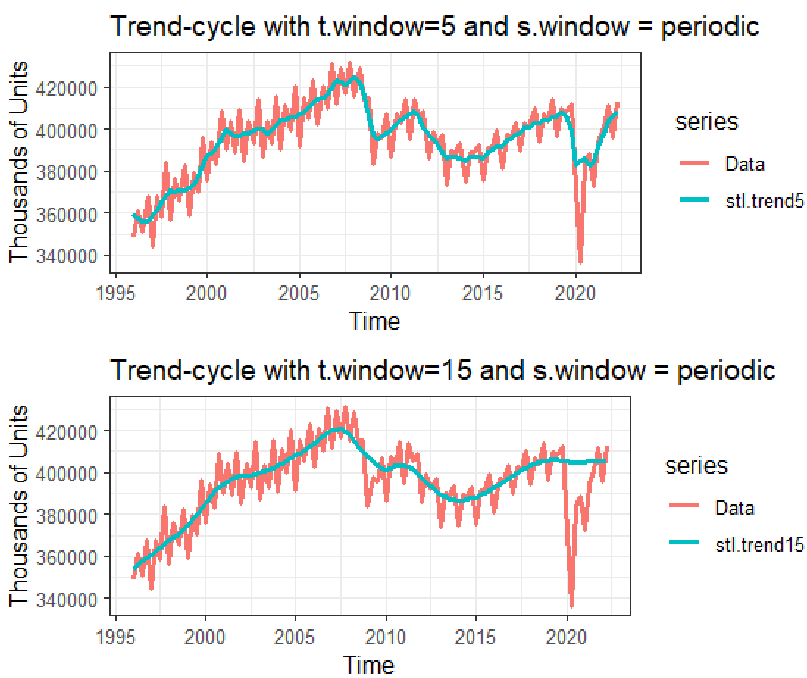

Figure 9: **Trend-cycle with t.windows(5) and t.windows(15)**

Figure 9 show different values of t.windows with s.windows equal to periodic. With t.window equal to 5 we have a less smooth trend, while with a value of t.window equal 15 we have a very smooth trend.

In order to have a good balance between overfitting the seasonality and the possibility of this changing slowly over time, we use the function mstl() in which the parameter t.window and s.window is chosen automatically. Overfitted model to data can lead to very wrong predictions.

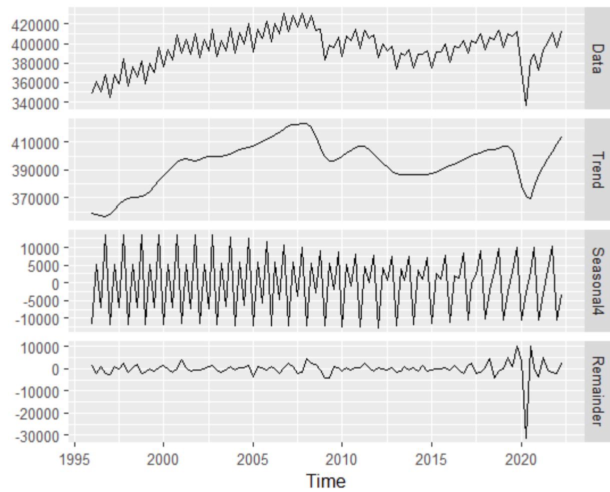

Figure 10: **STL decomposition with mstl() function**

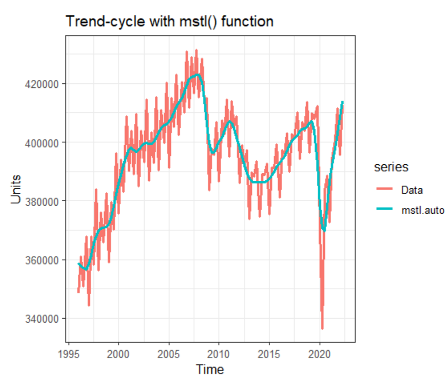

Figure 11: **Trend-cycle with mstl() function**

The figure 10 shows the automated STL decomposition, while figure 11 shows the trend-cycle component.

3 ARIMA Models

3.1 Preliminary analysis

For the construction of an Arima model, several steps must be followed; first, a preliminary analysis must be done. The fundamental assumption in time series analysis is that the time series is a realization of a stationary process, so the first thing to do is to determine whether the time series can be considered a realization of a stationary process; if not, a transformation must be performed.

A time series can be considered a realization of a stationary process if: there is no systematic variation in the mean (no trend) and variance, and there is no periodic variation.

There are two types of nonstationarity, heteroschedasticity (non-constant variance) and nonstationarity of the mean (trend). To obtain homoschedasticity we use the Box-Cox transformation. We need to find the optimal value of λ^* so that the series is as homoschedastic as possible. The optimal value of λ^* was found to be 2. A power transformation was applied to the original data:

$$y_t = \frac{x_t^2 - 1}{2} \tag{1}$$

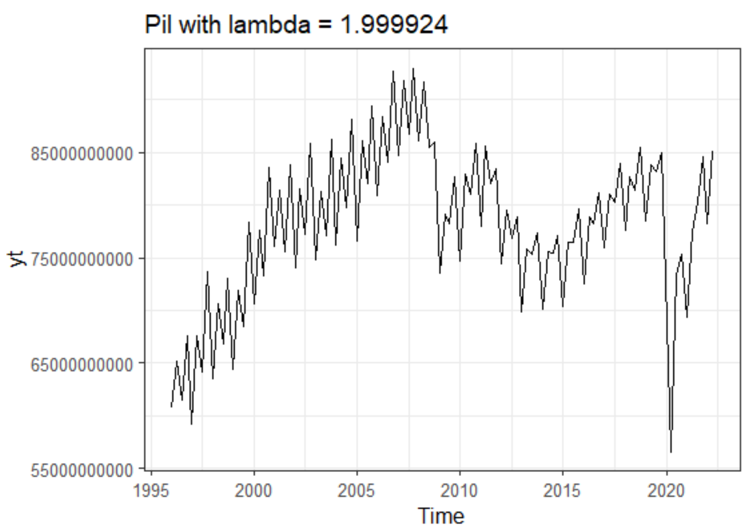

Figure 12: **PIL transformation with λ = 2**

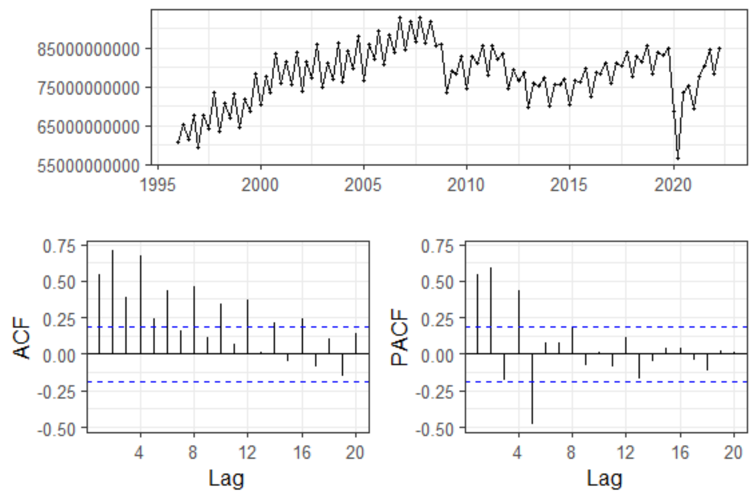

Figure 13: **Acf and Pacf of the PIL with $\lambda = 2$**

Now turn to nonstationarity in mean. As can be seen from Figure 13, the slow decline of the ACF toward zero suggests to us that a trend is present; in addition, we can also note the presence of seasonality.

When the series has seasonality of period s, we can remove seasonality from the time series. We can remove seasonality by differentiating it (variations from one season to the next):

$$\Delta_s X_t = (1 - L^s)X_t = X_t - X_{t-s} \tag{2}$$

In this case we have quarterly data, so there are 4 periods

$$\Delta_4 X_t = (1 - L^4)X_t = X_t - X_4 \tag{3}$$

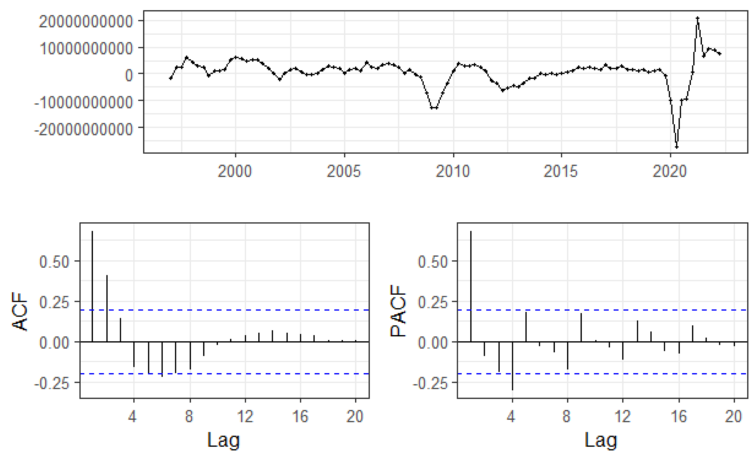

Figure 14: **Acf and Pacf of the seasonal differentiating with s = 4**

As we can see from Figure 14, the ACF shows that the global autocorrelations for the first lags decay very slowly to 0, this tells us that there is still some trend in the data that needs to be removed. In this case non-seasonal and seasonal differences can be combined to stationarize the series:

$$\Delta^d \Delta_s^D X_t = (1-L)^d (1-L^s)^D X_t \tag{4}$$

where:

- D should never be more than 1
- d+D should never be more than 2

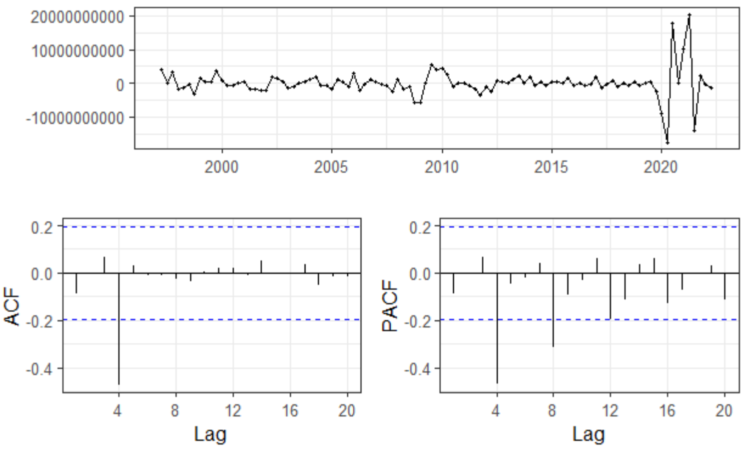

Figure 15: **Acf and Pacf of the combined non-seasonal and seasonal differences**

As we can see from figure 15, the time series obtained combined non seasonal and seasonal differences exhibit the typical pattern of a stationary time series. Graphically, stationary series are approximately horizontal, with constant variance. The ACF shows that the global autocorrelations decay very quickly to 0. Another empirical criterion for the choice of d is to determine the minimum value for which:

$$Var(\Delta^d X_t) < Var(\Delta^{d+1} X_t) \tag{5}$$

Table 2 shows the variance and the percentage reduction of the variance considering the various transformations carried out on the time series.

Table 2: **Variance and reduction in variance %**

Series	**Variance**	**Reduction in variance %**
y_t	54829114531582042112	-
$\Delta_4 y_t$	27552108509915820032	0.4974913
$\Delta\Delta_4 y_t$	17373057753860663296	0.6831417

Another criterion for verifying stationarity in a time series is to carry out a test. Augmented Dickey Fuller test (ADF Test) is a common statistical test used to test whether a given time series is stationary or not. It is one of the most commonly used statistical test when it comes to analyzing the stationary of a series.

Table 3: **Augmented Dickey-Fuller Test (alternative hypothesis: stationary)**

Series	Dickey-Fuller	Lag order	p-value
$\Delta\Delta_4 y_t$	-5.2183	4	0.01

As we can see from Table 3, $\Delta\Delta_4 y_t$ is a stationary series. The p-value (0.01) is less than 0.05, so the null hypothesis H0 (non-stationary) is rejected.

3.2 Model Identification

Model identification involves determining the model order required to capture the salient dynamic features of the data; the model order (p,q) can be determined using the ACF and PACF graph.

For the seasonal ARIMA models the seasonal part of an AR or MA model will be seen in the seasonal lags of the PACF and ACF. The modeling procedure is almost the same as for non-seasonal data, except that we need to select seasonal AR and MA terms as well as the non-seasonal components of the model.

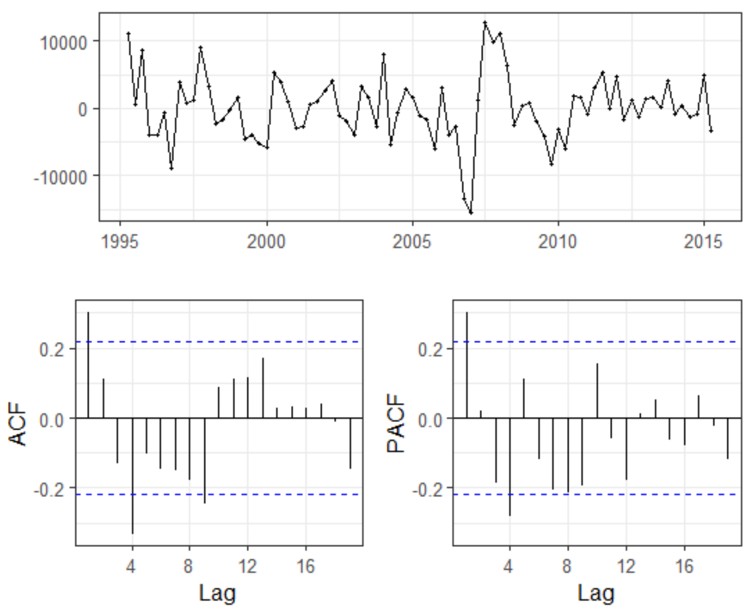

Figure 16: **Acf and Pacf of the combined non seasonal and seasonal differences on the Training set (Qtr1 1996—Qtr4 2017)**

In figure 16 we analyze the ACF and PACF of the $\Delta\Delta_4 y_t$ stationary series. Our aim now is to find an appropriate ARIMA model based on the ACF and PACF.

The significant spike at lag 1 in the ACF suggests a non-seasonal MA(1) component, and the significant spike at lag 4 in the ACF suggests a seasonal MA(1) component.

In PACF there seems to be an exponential decay of partial autocorrelations (lag 4, 8, ...),furthermore, the significant peak at lag 1 indicates a nonseasonal AR(1) component and the peak at lag 4 suggests a seasonal AR(1) component.

Consequently, this initial analysis suggests that a possible model for these data is an $ARIMA(1,1,1)(1,1,1)_4$. We fit this model, along with some variations on it, and compute the various information criteria shown in the table 4

Table 4: **Arima models with information criteria**

Model	**AICc**	**BIC**	**HQIC**
$ARIMA(1,1,1)(1,1,1)_4$	3771.35	3782.66	3772.46
$ARIMA(1,1,1)(0,1,1)_4$	3769.09	3778.25	3769.49
$ARIMA(1,1,1)(1,1,0)_4$	3772.30	3781.46	3772.70
$ARIMA(1,1,0)(1,1,1)_4$	3769.36	3778.53	3769.76
$ARIMA(0,1,1)(1,1,1)_4$	3770.45	3779.62	3770.85

Among these models, the best is $ARIMA(1,1,1)(0,1,1)_4$ as it has the lowest AICc, BIC and HQIC values. The corrected AIC (Akaike's Information Criterion) is useful for determining the order of an ARIMA model. It can be written as:

- $AICc = AIC + \frac{2(p+q+k+1)(p+q+k+2)}{T-p-q-k-2}$

Where AIC can be written as:

- $AIC = -2log(L_{max}) + 2(p+q+k+1)$

$log(L_{max})$ is the maximized value of the log-Likelihood for the estimated model, while the last term in parentheses is the number of parameters in the model.

Another information criteria is the Bayesian Information Criterion (BIC). It can be written as:

- $BIC = AIC + [log(T) - 2](p+q+k+1)$

The difference between BIC and AIC manifests itself when we add a very large number of parameters in order to increase the goodness of fit of the model. In this case, the BIC penalizes this increase in parameters more (compared to the AIC).

Finally, the last information criterion is the Hannan-Quinn information criterion (HQIC). It can be written as:

- $HQIC = -2log(L_{max}) + 2(p+q+k+1)\ log(log(T))$

$log(L_{max})$ is the maximized value of the log-Likelihood for the estimated model, while the second term in parentheses is the number of parameters in the model. T is the number of observations. We obtain good models by minimizing the above information criteria.

We now turn to the analysis of the residuals of the best model chosen based on the information criteria. The goal is to check the adequacy of the models; the residuals must be small and must not exhibit systematic or predictable patterns.

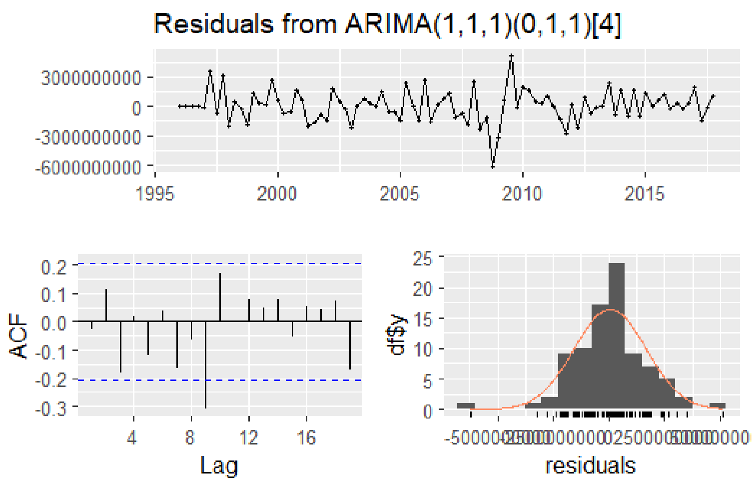

Figure 17: **Check residuals for ARIMA (1,1,1) (0,1,1)$_4$**

As can be seen from Figure 17, only one autocorrelation of the residuals falls outside the 95% confidence interval.

A very useful statistical test is the Ljung Box test , a type for testing whether a group of autocorrelations of a time series is non-zero; the test checks "overall" randomness based on a certain number of lags, and is thus a portmanteau test. Table 5 shows the results of the Ljung Box test.

Q*	Total lags used	p-value
9.0173	8	0.1084

Table 5: **Ljung Box test for ARIMA (1,1,1) (0,1,1)$_4$**

The p-value (0.1084) is greater than 0.05, so the null hypothesis H0 is accepted, this means that up to lag 8 all the global autocorrelations (acf) of the residuals are equal to 0, what has been said we can also see from the ACF shown in figure 17.

To understand whether the residuals have a normal distribution, we analyze the Q-Q plot and perform a test for normality, such as the Jarque-Bera test.

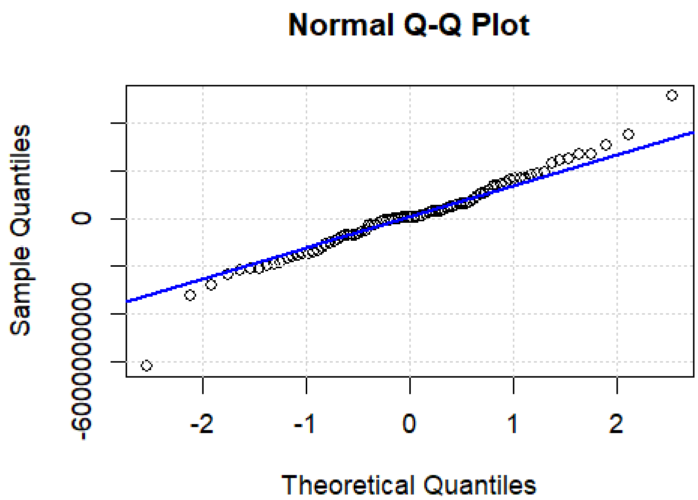

Figure 18: **Normal QQ PLOT**

In Figure 18 we can see the normal Q-Q plot. If all the points plotted on the graph lie perfectly on a straight line (y = x), then we can say that the distribution is of Normal type because it is uniformly aligned with the standard normal variable. In this case, the residuals do not have a normal distribution.

The Jarque-Bera test is a test of whether the sample data have skewness and kurtosis corresponding to a normal distribution. The null hypothesis is the joint hypothesis that skewness is zero and excess kurtosis is zero. Samples of a normal distribution have a skewness of 0 and a kurtosis excess of 0 (which is equivalent to a kurtosis of 3). Table 6 shows the results of the Jarque-Bera test.

X-squared	**p-value**
17.876	0.0001313

Table 6: **Jarque-Bera test**

The p-value (0.0001313) is less than 0.05, so the null hypothesis H0 is rejected. The residuals do not have a normal distribution.

3.3 Forecasting ARIMA models

Now, we will compare some models fitted so far using a test set consisting of the last five years of data. Thus, we fit the models using data from 1996 to 2017, and forecast PIL for the period from 2018 to 2022. The results are summarized in Table 7.

Table 7: **Arima models with information criteria**

Model	RMSE	MAE	MAPE	MASE
$ARIMA(1,1,1)(1,1,1)_4$	29648.05	21537.28	5.71	3.10
$ARIMA(1,1,1)(0,1,1)_4$	29594.26	21480.88	5.68	3.08
$ARIMA(1,1,1)(1,1,0)_4$	31682.44	23694.86	6.26	3.41
$ARIMA(1,1,0)(1,1,1)_4$	29721.80	21619.20	5.73	3.10
$ARIMA(0,1,1)(1,1,1)_4$	28955.75	20743.40	5.50	2.99

All accuracy measures in table 7 choose the $ARIMA(0,1,1)(1,1,1)_4$ model which has the lowest RMSE, MAE, MAPE and MASE value on the test set.

The two most commonly used scale-dependent measures are based on the absolute errors or squared errors:

- Mean absolute error: MAE $= mean\ (|\hat{e}_t|)$
- Root mean squared error: RMSE $= \sqrt{mean\ (\hat{e}_t^2)}$

Mape is a measure of accuracy based on the concept of percentage errors. Percentage errors have the advantage of being unit-free, and so are frequently used to compare forecast performances between data sets.

Measures based on percentage errors have the disadvantage of being infinite or undefined if $x_t = 0$ for any t in the period of interest, and having extreme values if any x_t is close to zero. The percentage error is given by $p_t = 100\ \frac{\hat{e}_t}{x_t}$.

- Mean absolute percentage error: MAPE $= mean\ (|p_t|)$

An alternative is MASE. The mean absolute scaled error:

- Mean absolute scaled error: MASE $= mean\ (|q_j|)$

Where:

$$q_j = \frac{e_j}{\frac{1}{T-1}\sum_{t=2}^{T}|x_t - x_{t-1}|}$$

For a non-seasonal time series, and:

$$q_j = \frac{e_j}{\frac{1}{T-m}\sum_{t=m+1}^{T}|x_t - x_{t-m}|}$$

For seasonal time series.

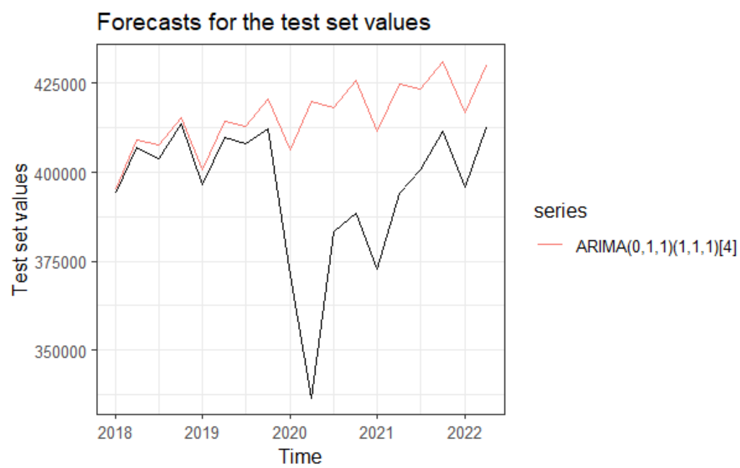

Figure 19: **Forecasts on the Test set**

Figure 19 shows the $ARIMA(0,1,1)(1,1,1)_4$ model that minimized the various accuracy measures seen in table 7. This model manage to follow the true trend of test set data quite closely.

The predictions obtained in Figure 19 and the accuracy measures obtained in Table 9 refer to the original scale. After choosing a transformation, we need to predict using the transformed data. Then, we need to invert the transformation (back-transform) to obtain predictions on the original scale. The inverse Box-Cox transformation in this case is given by:

$$x_t = \left\{(2y_t + 1)^{\frac{1}{2}}\right. \tag{6}$$

Where 2 is the lambda parameter and y_t is the time series to which the Box-Cox transformation has been applied. However, the point forecast from the "back-transformation" can be considered the median and not the mean of the distribution of the forecasts (assuming that the distribution over the transformed space is symmetric), but in some cases an "average forecast" is required. The back-transformed mean in this case is given by:

$$x_t = \left\{(2y_t + 1)^{\frac{1}{2}} \left[1 + \frac{\sigma_h^2(1-2)}{2(2y_t+1)^2}\right]\right. \tag{7}$$

Where σ_h^2 is the h-step forecast variance. The larger the forecast variance, the bigger the difference between the mean and the median. The difference between the simple back-transformed forecast and the back-transformed mean is called the bias.

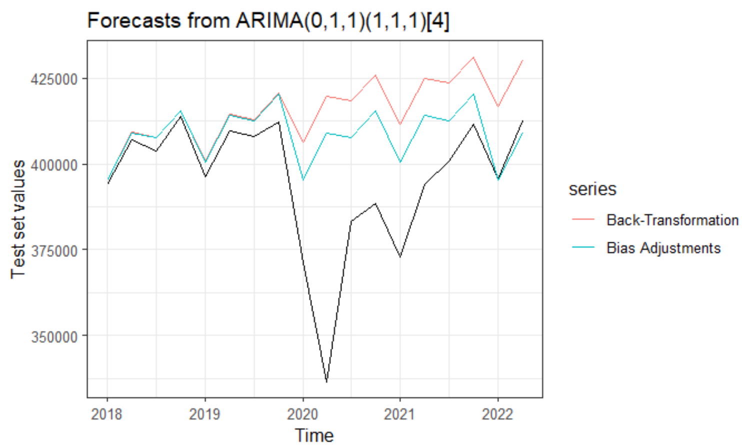

Figure 20: **Forecast medians and forecast means on the test set with $ARIMA(0,1,1)(1,1,1)_4$ model**

Figure 21 shows the difference between the forecast medians and the forecast means on the test set with $ARIMA(0,1,1)(1,1,1)_4$ model. From 2018 to 2020 the Back-Transform and Bias Adjustments forecasts are very similar, from 2020 the Bias Adjustments forecasts are better.

Finally, we use the $ARIMA(0,1,1)(1,1,1)_4$ model to forecast PIL in the last two quarters of 2022 and the four quarters of 2023 and 2024.

Figure 21 shows the point forecast (Bias-Adjustments) for the possible PIL in the period 2022-2024. In addition to the point forecast, the 95% and 80% confidence intervals are also shown.

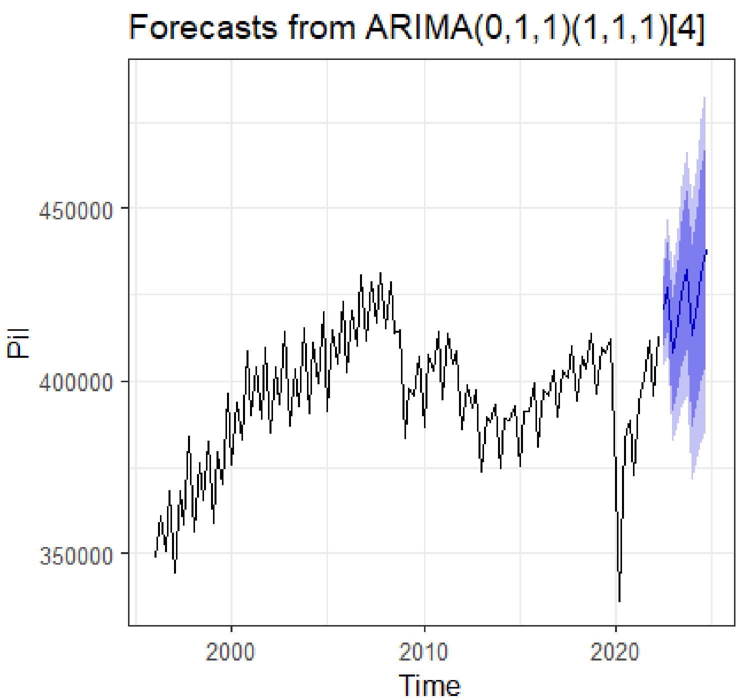

Figure 21: **Forecasts**