2022

Expenditure Incurred by International Tourists in Italy Following the COVID-19 Pandemic

CAPOZIO SARA | S.CAPOZIO@CAMPUS.UNIMIB.IT

1. INTRODUCTION

The tourism sector was among the industries most affected by the COVID-19 pandemic. The restriction of travel and the closure of tourism-related activities led to a significant decline in tourist flows in Italy.
The objective of this analysis is to assess whether the impact of the pandemic on the expenditure incurred by international tourists differed among the following territorial areas: North-East, North-West, Center, South, and Islands, and whether this effect was reabsorbed in the years following the global emergency.
The dataset under examination includes three time series covering a period of 26 years with quarterly frequency. The data spans from the first quarter of 1997 to the fourth quarter of 2022.
The variable under analysis is "Expenditure by territorial area", which includes all travel-related expenses incurred before departure, during the trip, or upon return. This expenditure comprises costs for transportation, accommodation, meals, various purchases, and recreational, cultural, and sports activities. The scale of this variable is expressed in millions of euros.
In the first part of the analysis, two control variables are introduced:

- "Travelers at destination": the number of travelers visiting specific locations. The scale of this variable is expressed in thousands.
- "Overnight stays by territorial area": the number of nights spent by guests in accommodation facilities during the considered period. The scale of this variable is expressed in thousands.

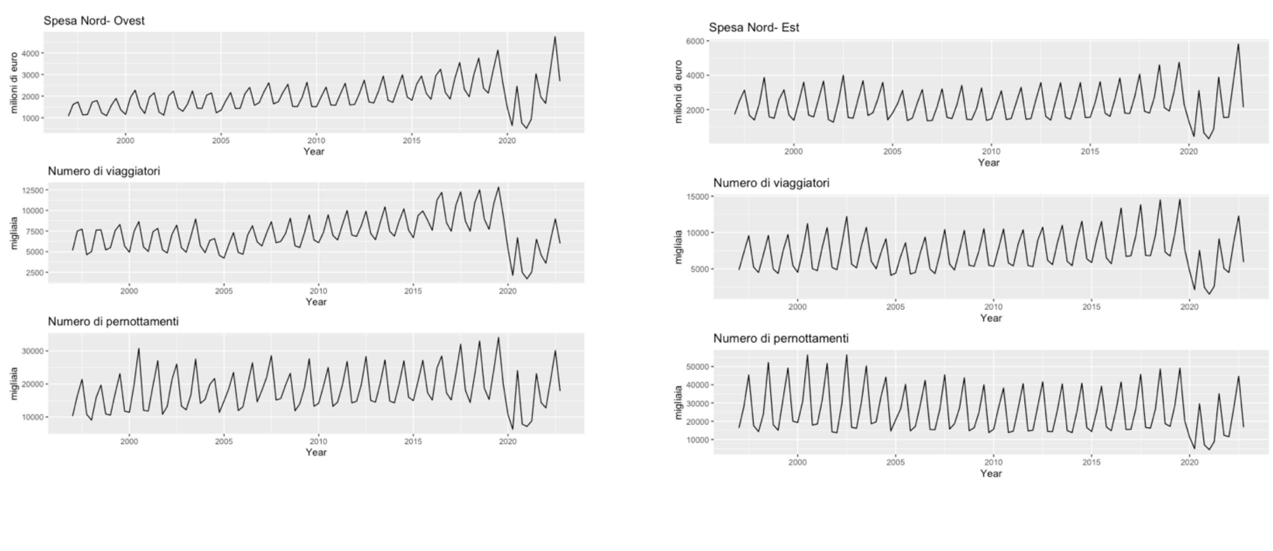
These two variables are used for interpretative purposes to better understand spending behavior over the time period of interest.
A second objective of the analysis, which will be explored in detail in the regression chapter, is to identify the main determinants that may have contributed to an increase or decrease in total expenditure by tourists in Italy. The economic covariates analyzed are:

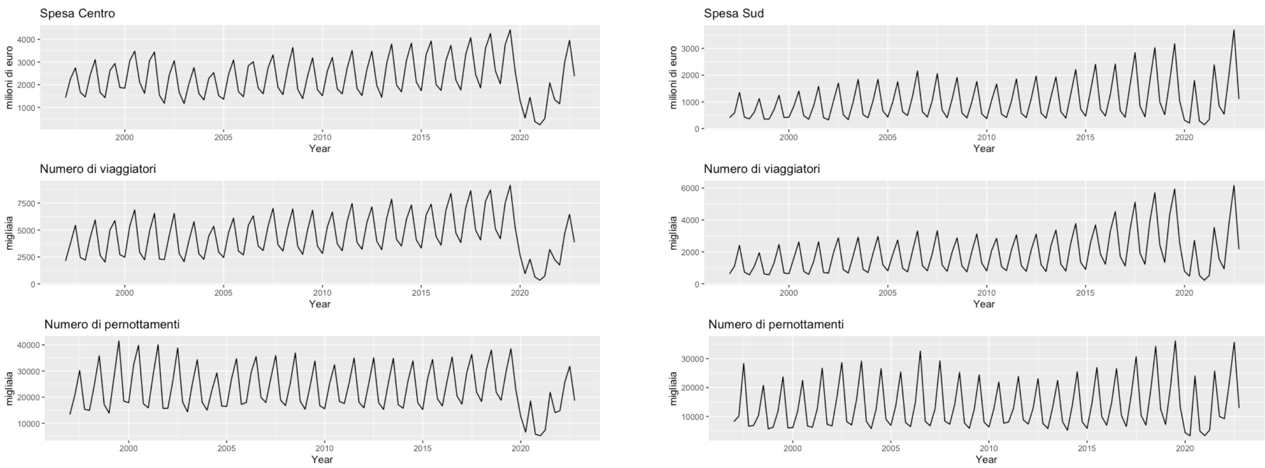
- "Consumer price index for the entire population": a statistical measure obtained through the average prices of a set of goods and services, generally used as an indicator of inflation. This variable is expressed in millions of euros.
- "Debt": the accounting balance between revenues and expenditures. This variable is expressed in millions of euros.
- "Gross fixed investments": purchases of durable material goods made by a company during the fiscal year, including the acquisition of machinery, equipment, furniture, transportation means, buildings, land, and the increase of fixed capital for internal work. The scale of this variable is expressed in millions of euros.
- "Employee income": wages received in the context of an employment relationship. The scale of this variable is expressed in millions of euros.

2. GRAPHICAL ANALYSIS OF TIME SERIES

By graphically analyzing the three time series under examination for each macro-region, it is possible to preliminarily identify any unobservable components, specific patterns, or potential structural breaks.

Figure 1: Graphical Representation of Time Series





From Figure 1, we can observe the presence of a strong seasonal component and a sudden temporary variation (structural break) that occurred in the first quarter of 2020 due to the COVID-19 pandemic.

Regarding the temporal trend of the variables, the time series characterizing expenditure and travelers differ by territorial area.

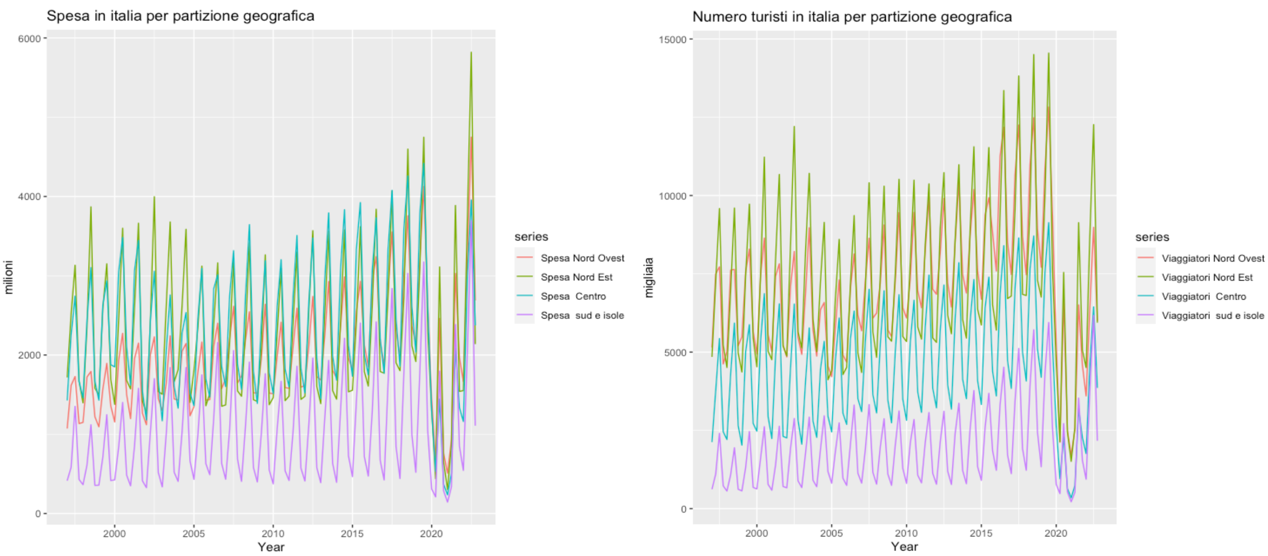
In the North-West and Central Italy, the expenditure incurred by foreign tourists follows a linear increasing trend, whereas in the North-East, it appears to follow a slightly parabolic pattern. In Southern Italy, the phenomenon appears heteroscedastic, with variability directly proportional to the time dimension; however, the data does not follow a specific functional form.

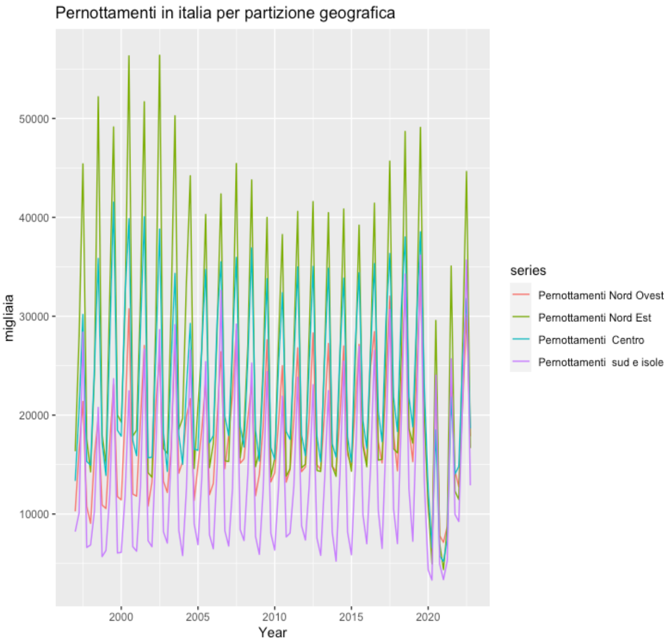
The time series related to the number of travelers in both the northern and central areas of the country shows a significant change in slope recorded in 2005, leading to a continuous increase in the number of tourists in these regions. In Southern Italy, however, a constant increase in the number of tourists has been observed only since 2015.

Despite the increase in the number of travelers over time, the number of overnight stays has remained constant in all areas except for the North-West, showing no significant increases or decreases.

To identify similarities or differences between geographic areas and assess the impact of the pandemic, the time series are overlaid.

Figure 2: Graphical Representation of Overlaid Time Series





Considering the time window before the onset of the pandemic, the North-East and North-West regions attracted the majority of international tourist flows. Additionally, the expenditure of Italian tourists is significantly concentrated geographically, with the North-East and Central Italy accounting for more than half of the total spending. From the graphs, we can observe that the Central Italy region, despite recording a lower number of visitors compared to the North-West, has a significantly higher average total expenditure, amounting to 322.9 million euros, making it one of the most expensive destinations for foreign tourists.Furthermore, there is a clear discrepancy between international traveler flows and the tourism potential of Southern Italy. The latter represents the region with the lowest number of visitors, absorbing only 10% of the total expenditure. In 2020, with the outbreak of the SARS COVID-19 pandemic, the total expenditure of foreign travelers in Italy decreased by 17.3 billion euros. From the graphs, we can observe that the northern and central regions were the most affected by the pandemic. It is evident that the impact of COVID-19 on tourism has been absorbed across all areas, except for Central Italy, which remains the most affected region. The tourism sector in this area is still in the recovery phase, as in the fourth quarter of 2022, the number of tourists and total expenditure remained lower than in the fourth quarter of 2019. Southern Italy, on the other hand, has experienced a positive recovery, recording an increase in both the number of visitors and overnight stays in 2022.

3. BOX-JENKINS PROCEDURE:

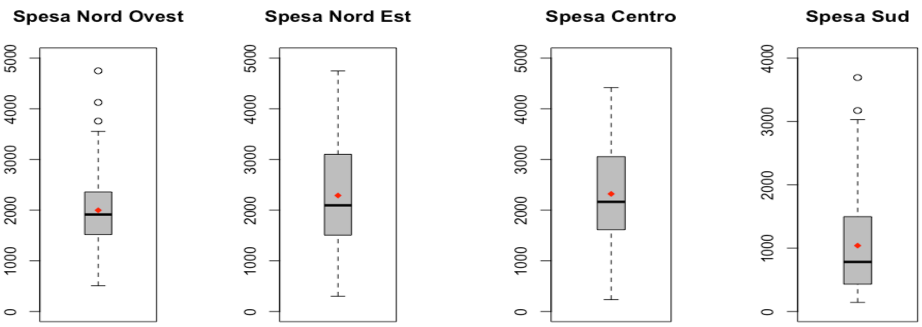
Using the Box & Jenkins methodology, it is possible to identify the optimal model capable of best a pproximating the temporal trend of expenditure by geographic area. The procedure consists of the following phases:

1. Exploratory analysis
2. Data transformation
3. Stationarity analysis
4. Analysis of time series components
5. Model estimation and evaluation of fit performance

3.1 EXPLORATORY ANALYSIS

Exploratory analysis allows for the identification of the type of distribution of the variables under examination and the possible existence of outliers.

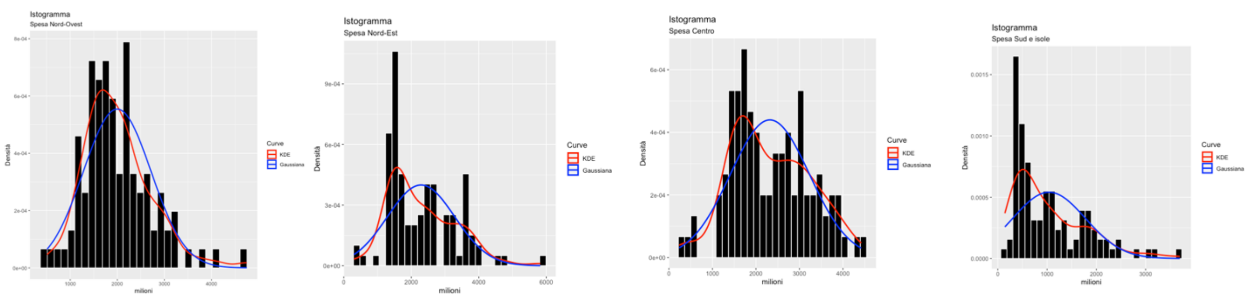
Figure3: box-plot



By analyzing the box plots, the examined variables are all characterized by positive skewness, as the mean is significantly higher than the median. Additionally, the presence of anomalous values can be observed,

deviating from the expenditure distribution of the North-West and South macro-regions.
A detailed analysis of the outlier values shows that they correspond to a sharp increase in tourist expenditure in the North-Western area during the third quarter of 2019 and 2022. Specifically, an increase of 10% and 30% was recorded compared to 2018.
Similarly, in Southern Italy, the third quarter of 2019 and 2022 showed an increase in expenditure of 5% and 20%, respectively, compared to 2018.
The analysis of the histograms further confirms the positive skewness hypothesis of the data, as the empirical density distributions for each variable in every macro-region deviate from the theoretical normal distribution.

Figure 4: Histograms



The evidence obtained through the graphical analysis appears to be consistent with the results generated by the Bera-Jarque and Shapiro-Wilk normality tests.

Table 1: Bera-Jarque Test

	Test Bera-Jarque			
SPESA	**Nord-Ovest**	**Nord-Est**	**Centro**	**Sud**
p-value	2.294e-06	0.00308	0.3717	1.461e-07

Observing Table 1, it can be noted that, for every level of significance, there is sufficient empirical evidence to reject the null hypothesis that the distributions have a skewness index of zero and a kurtosis index of three. Only expenditure in Central Italy appears to exhibit a symmetric distribution.

Table 2: Shapiro-Wilk Test

	Test Shapiro-Wilk			
SPESA	**Nord-Ovest**	**Nord-Est**	**Centro**	**Sud**
p-value	0.001253	4.106e-05	0.04524	3.326e-08

The Shapiro-Wilk test, based on the correlation between empirical data and theoretical normality values, rejects the null hypothesis of data normality at every level of significance.
Although expenditure in Central Italy is symmetric, its distribution does not follow the pattern of a normal distribution.
This result can be inferred by observing the deviation between the theoretical and empirical density functions, as represented in the histograms

3.2 DATA TRANFORMATION:

A useful tool for addressing the issue of non-normality in the data and the potential presence of heteroscedasticity is the Box-Cox transformation.
The objective of this transformation is to identify a variance-stabilizing function, ensuring that the observations composing the time series exhibit constant variability over time.

Figure 5: Box-Cox Graphs

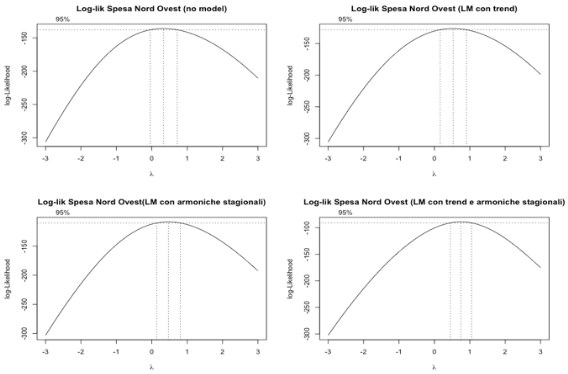
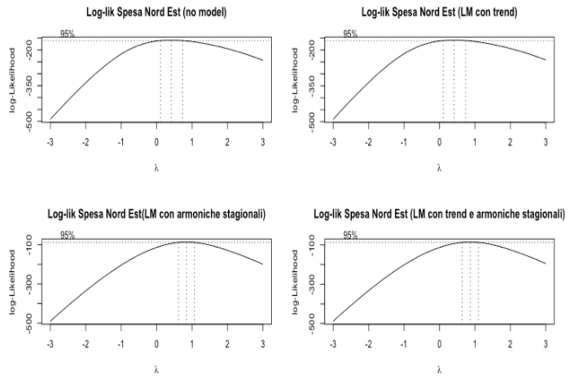
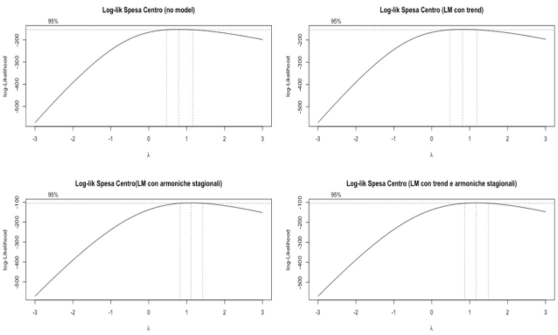


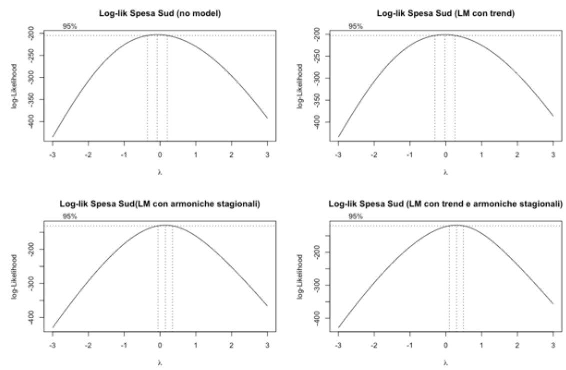
Figure 5 represents the optimal lambda values for expenditure in the North-Western region, obtained by applying Guerrero's method and the profile log-likelihood maximization method across different models, where the target variable is regressed as a function of trend, seasonality, and their combination.
Observing the graph related to the total expenditure incurred by tourists, all four 95% confidence intervals include the value ½, suggesting the application of the square root transformation to the variable of interest.



Regarding expenditure in the North-Eastern area, the confidence intervals appear to be highly variable. Therefore, the decision is made to focus on the fourth model, which is built on seasonally adjusted and detrended data, as removing the respective components allows for capturing the underlying trend of the variables.
Since the confidence interval includes the unit value, it is decided not to apply any transformation.

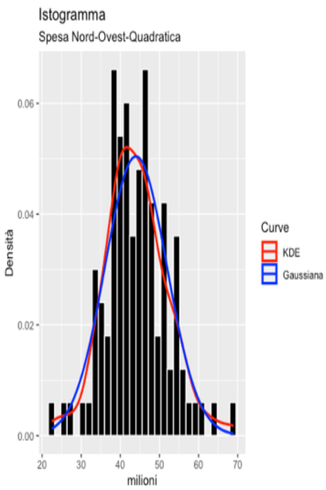


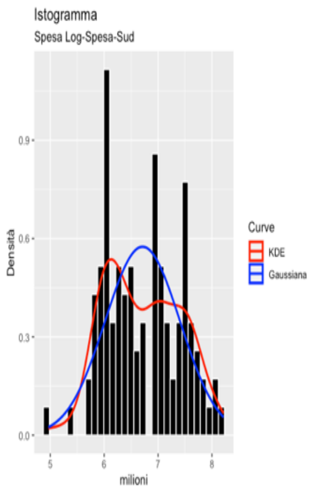
Analyzing expenditure in Central Italy, the confidence intervals constructed around the optimal lambda value are consistent with each other. Since they include the unit value, the variable does not require any transformation.



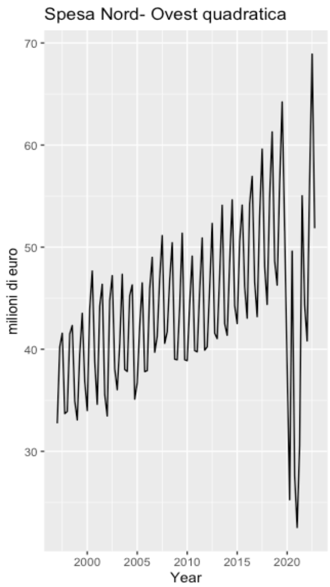
Finally, considering the Southern area, we can observe that the confidence intervals constructed around the optimal lambda parameter for the ‘Expenditure’ variable include zero, suggesting the application of a logarithmic transformation to the data.

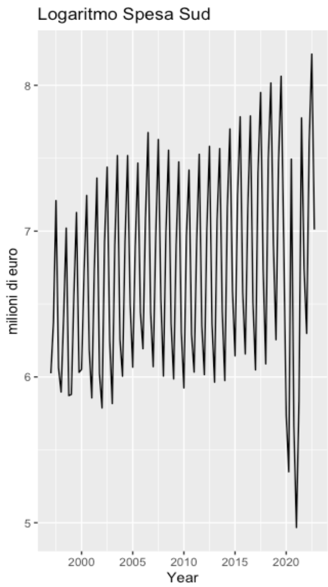
Once the appropriate transformations have been identified, we proceed with analyzing the distributions and graphs of the transformed variables.





Observing the histogram, the transformation applied to the ‘North-West Expenditure’ variable appears to have improved the skewness of its distribution, making it normal.
This result is confirmed by the p-value of 0.2314, obtained using the Bera-Jarque test.
Regarding the logarithmic transformation applied to the Southern Expenditure variable, the transformed distribution appears to be symmetric, although its empirical density function does not perfectly match the theoretical normal distribution.
The Bera-Jarque test confirms the findings from the graphical analysis, yielding a p-value of 0.1749.





The square root and logarithmic transformations, when analyzing the temporal trend of the transformed variables, appear to have stabilized the seasonal fluctuations of the series before the break that occurred in 2020.

3.3 PERSISTENCE ANALYSIS

To analyze the stationarity of the time series, the Dickey-Fuller test is performed to verify the null hypothesis of the presence of a unit root, indicating the existence of a stochastic trend.
Let us assume that the time series X_t can be interpreted as an AR(p) process that moves around a deterministic component CD_t. The hypothesis testing takes on different configurations depending on the form of the deterministic component:

- $CD_t = 0$, meaning the deterministic component is null
- $CD_t = k$, meaning the deterministic component is a nonzero constant
- $CD_t = k + \delta t$, meaning the deterministic component is linearly dependent on time

Three different tests will be conducted: the first to test whether the process is stationary with zero mean, the second to evaluate stationarity around a constant, and the third to assess stationarity around a deterministic trend.
Table 3 presents the results obtained by applying the Dickey-Fuller test, assuming the deterministic component is null.
The test verifies the following hypothesis system:

$$\begin{cases} H0\text{: The process is a random walk without drift} \\ H1\text{: The process is a stationary process with zero mean.} \end{cases}$$

Table 3: Test 1 Dickey-Fuller

Test 1 Dickey-Fuller			
WESTERN EXPENDITURE:: **Test Statistic** = -0.653	**1 pct**	**5 pct**	**10 pct**
	tau1= -2.58	tau1= -1.95	tau1= -1.62
EASTERN EXPENDITURE: **Test Statistic** = -2.3456	**1 pct**	**5 pct**	**10 pct**
	tau1= -2.58	tau1= -1.95	tau1= -1.62
CENTER EXPENDITURE: **Test Statistic** = -2.2096	**1 pct**	**5 pct**	**10 pct**
	tau1= -2.58	tau1= -1.95	tau1= -1.62
SOUTH EXPENDITURE: **Test Statistic** = -0.5567	**1 pct**	**5 pct**	**10 pct**
	tau1= -2.58	tau1= -1.95	tau1= -1.62

Observing the results in the table, the test statistic values for the North-Western and Southern areas fall within the acceptance region of the null hypothesis at every level of significance. Conversely, for Central Italy and the North-East, the test statistic values are lower than the critical value corresponding to the 5% and 10% significance levels, but higher than that of 1%. Table 4 presents the results obtained by applying the Dickey-Fuller test, assuming the presence of a drift, meaning a nonzero deterministic component. The test verifies the following hypothesis system

$$\begin{cases} H0\text{: The process is a random walk} \\ H1\text{: The process is a stationary process with mean different to zero} \end{cases}$$

Table 4: Test2 Dickey-Fuller

Test 2 Dickey-Fuller				
NORTH WEST	**Tau2:** -6.996	**1 pct**	**5 pct**	**10 pct**
		-3.46	-2.88	-2.57
	Phi1: 24.4981	**1 pct**	**5 pct**	**10 pct**
		6.52	4.63	3.81
NORTH EAST	**Tau2:** -12.7435	**1 pct**	**5 pct**	**10 pct**
		-3.46	-2.88	-2.57
	Phi1: 81.2004	**1 pct**	**5 pct**	**10 pct**
		6.52	4.63	3.81
CENTER	**Tau2:** -10.3593	**1 pct**	**5 pct**	**10 pct**
		-3.46	-2.88	-2.57
	Phi1: 53.6575	**1 pct**	**5 pct**	**10 pct**
		6.52	4.63	3.81
SOUTH	**Tau2:** -14.718	**1 pct**	**5 pct**	**10 pct**
		-3.46	-2.88	-2.57
	Phi1: 108.3201	**1 pct**	**5 pct**	**10 pct**
		6.52	4.63	3.81

Observing the values in the table, two test statistics can be noted. The tau2 test statistic verifies whether, in the autoregressive model, including a nonzero constant, there is a unit root. The phi1 test statistic tests the joint null hypothesis that the process is non-stationary and that the constant is equal to zero.
Since for all four variables, the test statistic values fall within the rejection regions at every level of significance, it is concluded that the process is stationary around a drift.
Finally, the case is considered where the deterministic component is a linear trend, leading to the following hypothesis system:

$$\begin{cases} H0\text{: The process is a random walk with drift} \\ H1\text{: The process is a stationary process } \textit{around a trend} \end{cases}$$

Table 5: Test 3 Dickey-Fuller

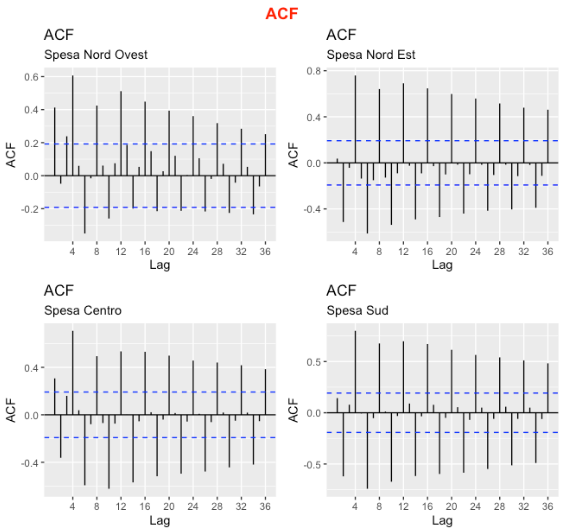
Test 3 Dickey-Fuller				
NORTH WEST	**Tau3:**	**1 pct**	**5 pct**	**10 pct**
	-8.5844 2	-3.99	-3.43	-3.13
	Phi2:	**1 pct**	**5 pct**	**10 pct**
	24.5914	6.22	4.75	4.07
	Phi3:	**1 pct**	**5 pct**	**10 pct**
	36.8571	8.43	6.49	5.47
NORTH EAST	**Tau3:**	**1 pct**	**5 pct**	**10 pct**
	-12.7519	-3.99	-3.43	-3.13
	Phi2:	**1 pct**	**5 pct**	**10 pct**
	54.2128	6.22	4.75	4.07
	Phi3:	**1 pct**	**5 pct**	**10 pct**
	81.3172	8.43	6.49	5.47
CENTER	**Tau3:**	**1 pct**	**5 pct**	**10 pct**
	-10.3467	-3.99	-3.43	-3.13
	Phi2:	**1 pct**	**5 pct**	**10 pct**
	35.6847	6.22	4.75	4.07
	Phi3:	**1 pct**	**5 pct**	**10 pct**
	53.5269	8.43	6.49	5.47
SOUTH	**Tau3:**	**1 pct**	**5 pct**	**10 pct**
	-16.0537	-3.99	-3.43	-3.13
	Phi2:	**1 pct**	**5 pct**	**10 pct**
	85.9153	6.22	4.75	4.07
	Phi3:	**1 pct**	**5 pct**	**10 pct**
	128.8608	8.43	6.49	5.47

Observing the values in the table, three test statistics can be noted.

- The tau3 test statistic verifies whether, in the autoregressive model including a linear trend, there is a unit root.
- The phi3 test statistic tests the joint null hypothesis that the process is non-stationary and that the slope of the trend is zero.
- The phi2 test statistic verifies that the process is non-stationary and that the parameters of the trend equation are jointly zero.

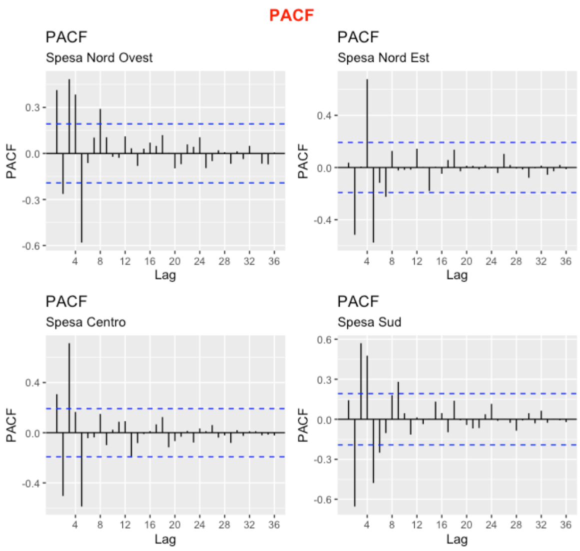
Once again, the test statistic values fall within the rejection regions at every level of significance, leading to the conclusion that the variables are stationary around a deterministic trend, a result already observed in the graphical analysis.
Having completed the stationarity analysis, we now proceed to examine the persistence of the time series using the autocorrelation function (ACF) and partial autocorrelation function (PACF), which show the variation in correlations between observations as the lags change.

Figure 6: ACF



Analyzing Figure 6, we observe that the correlations between lags are significantly different from zero at a 95% confidence level, as they extend beyond the confidence bands, represented by the dashed blue lines.
In all the graphs, the correlation at lag four is higher than at other lags. This is due to the seasonal component present in the data; in fact, peaks tend to repeat every four lags.
Additionally, as the lag increases, the autocorrelation tends to decrease, which is attributed to the trend component present in the data.

Figure7: PACF



Regarding the partial autocorrelation graph, after removing intermediate effects, it is observed that up to lag 5, the partial autocorrelations are significantly different from zero, while from lag 6 onward, they completely dissipate.

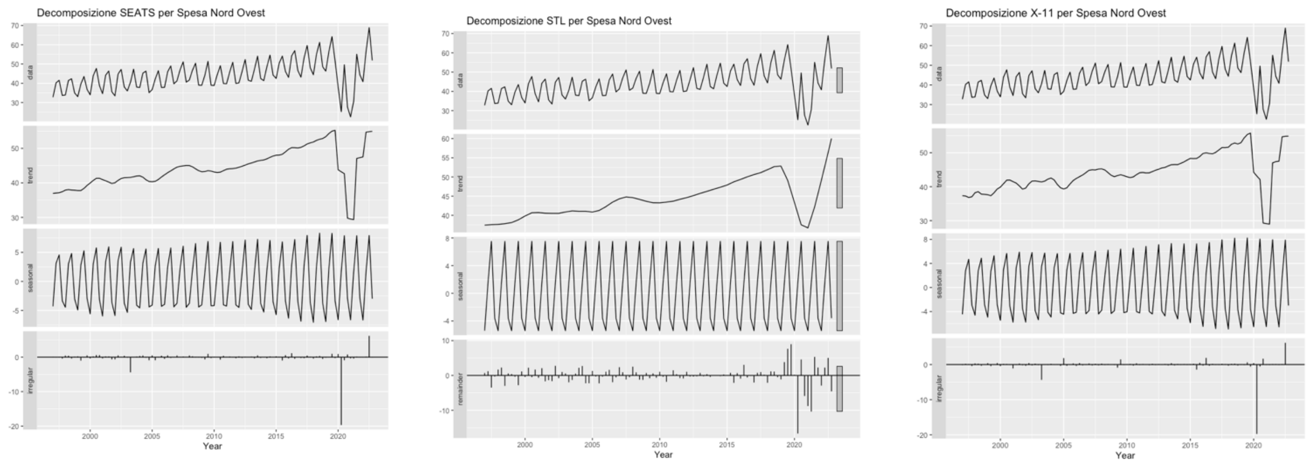
3.4 ANALYSIS OF TIME SERIES COMPONENTS

From the analyses conducted, the evolution of the variables over time appears to result from the aggregation of three unobservable components:

- Trend
- Seasonality
- Structural break

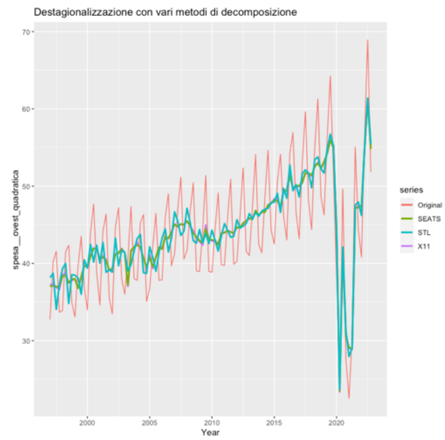
The examined variables are not heteroscedastic phenomena, as the magnitude of seasonal fluctuations does not change with increasing time span—an assertion confirmed by the Dickey-Fuller tests.
Thus, the relationship between the time series and their individual components appears to be additive.
The decomposition of the series was performed using methods robust to the presence of outliers and level shifts, specifically X-11, TRAMO-SEATS, and STL.

Figure8: Decompositions

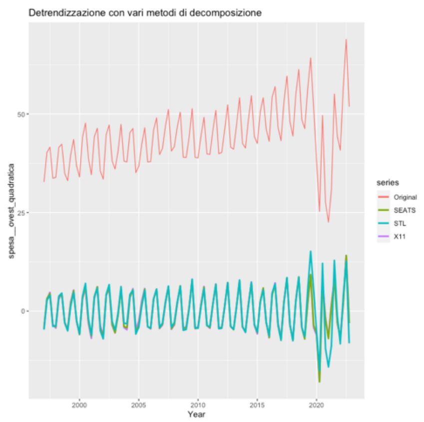


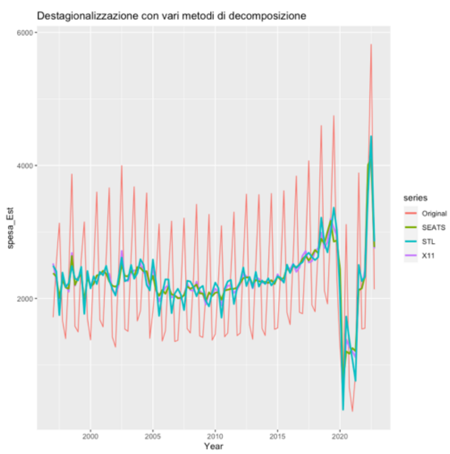
La Figura 8 evidenzia che tutti e tre i metodi di decomposizione catturano efficacemente la forte componente stagionale, il trend e la presenza del break strutturale.
Tuttavia, il metodo STL, a differenza degli altri due algoritmi, è in grado di individuare più accuratamente i valori outliers. Ciò è possibile grazie alla sua capacità di monitorare le componenti ciclo-trend e stagionalità attraverso un parametro di smoothing.
I metodi analizzati, pur rimuovendo il trend e la stagionalità, non riescono a isolare completamente il break strutturale, che quindi rimane una componente residua.

Figure9: Graph of seasonally adjusted and detrended series

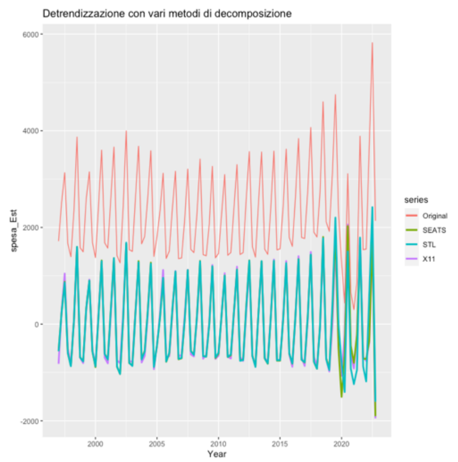


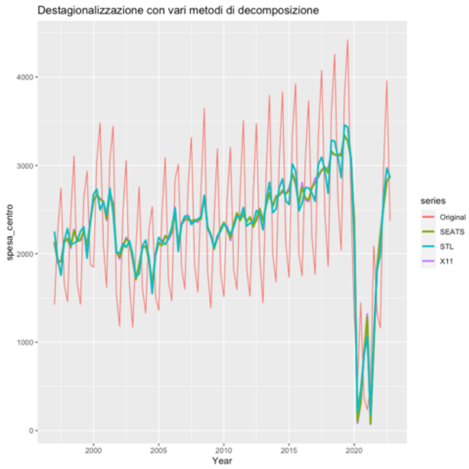
The overlap of the seasonally adjusted series for 'Expenditure' in the north-western region, shown in Figure 9, highlights the similarity between decomposition methods. Greater discrepancies emerge when analyzing the detrended series. As previously mentioned, only the STL method successfully captures the impact of shocks occurring from 2020 onwards.



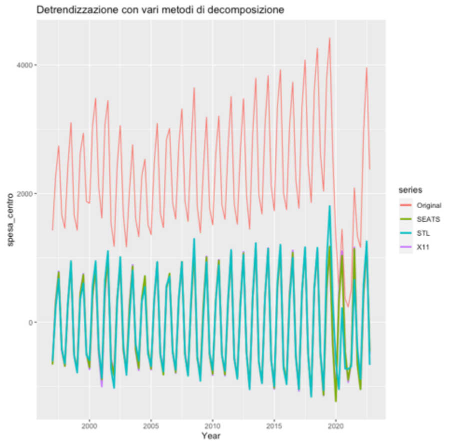


The considerations made for 'Expenditure' in the north-western region also apply to the other variables under analysis. In general, the seasonal adjustment process based on the smoothing parameter of the STL method makes the seasonally adjusted series less smooth compared to other methods, as the rate of variation in the seasonal component can be controlled by the user.

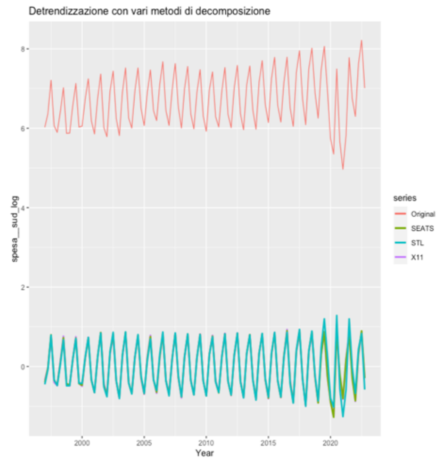




Regarding the detrending process, all three methods, by removing the trend, lose a significant portion of the information contained in the structural break component. The method that best captures the shocks is the STL method.

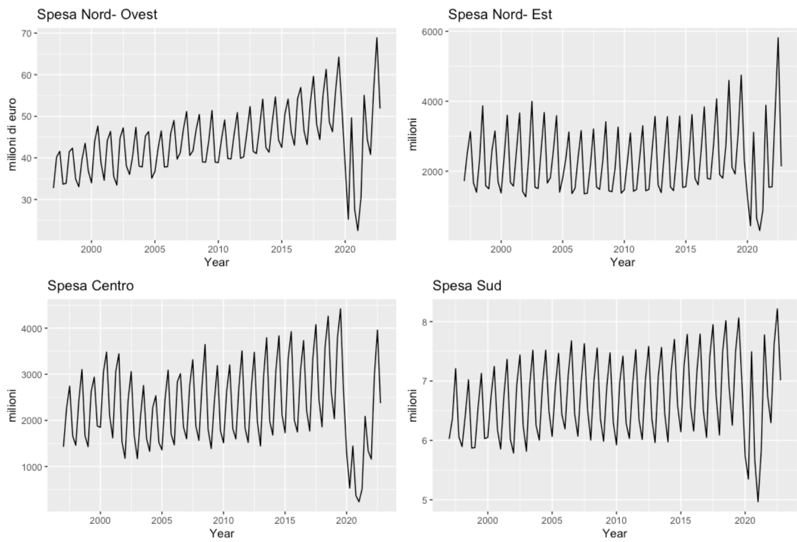


In the case of 'Expenditure' in Southern Italy and the islands, the decomposition methods appear to be equivalent.



3.5 SARIMA MODELS

Analyzing the temporal trend of tourist expenditure in Italy, two significant structural breaks can be observed. The first break coincides with the onset of the COVID-19 pandemic, which spread globally starting in December 2019. The shock caused by the first wave had a clear impact on expenditure variation in the following quarter, leading to a sudden level shift in the time series. The effect of the global emergency had an even more dramatic impact on winter tourism in 2021, due to the introduction of new restrictive measures following a rise in cases recorded in the summer of 2021.
With the declaration of the end of the state of emergency on March 31, 2022, the effects of the pandemic in Northeastern, Northwestern, and Southern Italy appear to have been reabsorbed, with expenditure levels returning approximately to those recorded in 2019. However, a different situation occurred in Central Italy, the area most affected by the pandemic's consequences and still in the process of economic recovery. In this region, the first quarter of 2022 recorded an expenditure level 40% lower than in 2019.



Starting from the second quarter of 2022, a strong increase in spending has affected, with varying intensity, all territorial areas, despite the fact that in recent years these areas have experienced lower tourist flows compared

to 2019. This sudden increase, in addition to being caused by the economic recovery following the pandemic, could be due to the impact of the outbreak of the war in Ukraine in February 2022 on the rise in commodity prices, including energy and fuel.
In eastern, western, and southern Italy, spending reached its highest level, increasing by 15%, 22%, and 16%, respectively, compared to the value recorded in the third quarter of 2019. This growth also affected, although not significantly, central Italy, where spending was still 10% lower than that observed in the previous three years.
To extract the two shocks, an intervention analysis was performed, which is a modeling procedure used to incorporate the effects of exogenous events or interventions on the series. In the first phase, the automated technique of Hyndman & Khandakar identifies the optimal SARIMA model for the pre-intervention (pre-COVID) sample, and in the second phase, this model is applied to all observations in the series. In this final step, two new regressors are added to the model.

- A first dummy dichotomous variable to identify the period of the state of emergency

$$X_{1t} = \begin{cases} 1 \; if \; 2020\,\mathrm{I}\,(\text{beginning of the pandemic}) \leq \mathrm{t} \leq 2022\,\mathrm{I} \\ (\text{declaration of the end of the state of emergency}) \\ 0 \; otherwise \end{cases}$$

- A second dummy variable to model the effect of the outbreak of the war in Ukraine in February 2022 on the increase in overall spending.

$$X_{2t} = \begin{cases} 1 \; if \quad 2022\, II \;(start\ of\ the\ war) \leq t \leq 2022\, IV (end\ of\ the\ war) \\ 0 \; otherwise \end{cases}$$

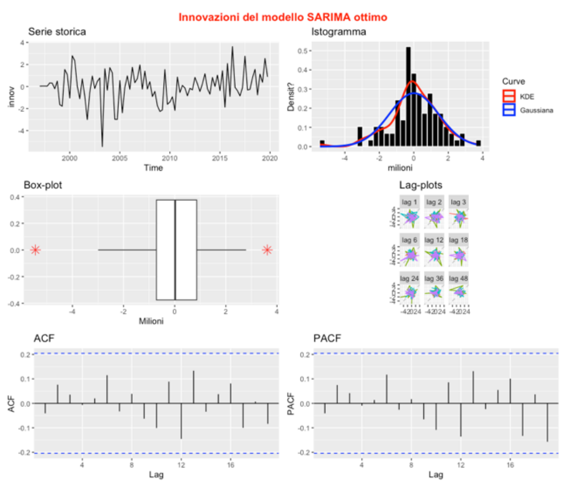
These dummy variables were applied with an ARMA filter in order to model the impact of the shocks on the level of spending and their speed of absorption in both the short and long term. A value of 1 was assigned to the first dummy variable for the entire pandemic period, not just at the time of its outbreak, because from February 2020 to March 2022, global government entities implemented numerous preventive measures that impacted the level of tourist spending in different ways. The same reasoning was applied to the dummy variable associated with the war, as the goal was to identify not only the direct effect of the contemporaneous shock caused by the outbreak of the conflict, but also all the effects it produced over time, such as rising energy prices, the increase in oil prices, and inflation.

3.5.1 SARIMA MODEL NORTH-WEST SPENDING

By applying the procedure outlined above to the spending recorded in north-western Italy, the following output is obtained:

Output1: Pre-intervention Model North-West Spending

```
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)
ar1  -0.36067    0.10122 -3.5632 0.0003663 ***
sma1 -0.52956    0.10474 -5.0560 4.281e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Coefficients:
          ar1     sma1
      -0.3607  -0.5296
s.e.   0.1012   0.1047

sigma^2 estimated as 2.508:  log likelihood = -164.16,  aic = 332.33
```

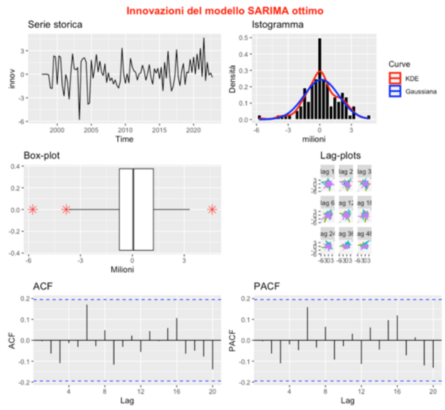


The optimal model for modeling the pre-intervention sample is an ARIMA (1,1,0) (0,1,1) model. From the model diagnostics, it can be observed that the residuals' behavior is very close to a random pattern; furthermore, there is no autocorrelation between the lags. Regarding their distribution, the residuals seem to follow a normal distribution, although the presence of an outlier on the left tail causes the distribution to be slightly negatively skewed.

Output2: Post-intervention Model North-West Spending

```
Coefficients:
          ar1     sma1  pulse-AR1  pulse-AR2  pulse-AR3  pulse-MA0  pulse-MA1  pulse-MA2  pulse-MA3  dummy1-MA0  dummy1-MA1
      -0.4466  -0.3908     0.3094    -0.8914    -0.3631   -13.1895   -17.8069     9.1863   -37.5673      6.3201    -23.9515
s.e.   0.1010   0.1166     0.0771     0.0606     0.1108     1.6250     2.0460     2.6784     3.3136      3.2630      3.4497
```

```
             Estimate Std. Error  z value  Pr(>|z|)
ar1         -0.446575   0.100971  -4.4228 9.742e-06 ***
sma1        -0.390767   0.116649  -3.3499 0.0008083 ***
pulse-AR1    0.309415   0.077144   4.0109 6.049e-05 ***
pulse-AR2   -0.891372   0.060647 -14.6978 < 2.2e-16 ***
pulse-AR3   -0.363098   0.110751  -3.2785 0.0010436 **
pulse-MA0  -13.189523   1.624984  -8.1167 4.790e-16 ***
pulse-MA1  -17.806912   2.046003  -8.7033 < 2.2e-16 ***
pulse-MA2    9.186283   2.678362   3.4298 0.0006040 ***
pulse-MA3  -37.567333   3.313576 -11.3374 < 2.2e-16 ***
dummy1-MA0   6.320122   3.263049   1.9369 0.0527605 .
dummy1-MA1 -23.951455   3.449734  -6.9430 3.839e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



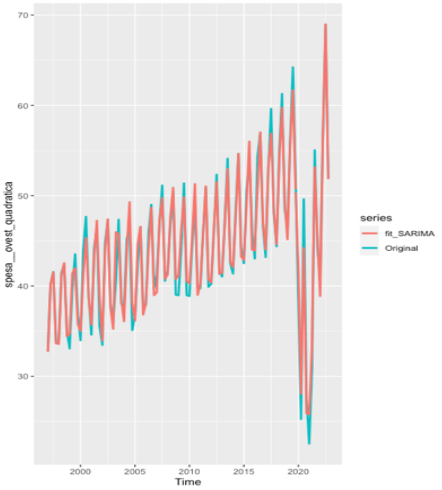
The identified dynamic model is as follows:

$$Y_t = w_0 X_{1t} + w_1 X_{1t-1} + w_2 X_{1t-2} + w_3 X_{1t-3} + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \delta_3 Y_{t-3} + \theta_0 X_{2t} + \theta_1 X_{2t-1}$$

where $X_{1t} = \begin{cases} 1 \ if \ 2020\, I \leq \ t \leq 2022\, I \\ 0 \ otherwise \end{cases}$ $\quad X_{2t} = \begin{cases} 1 \ if \ 2022\, II \leq \ t \leq 2022\, IV \\ 0 \ otherwise \end{cases}$

It can be observed that the effect of the pandemic is modeled by both short-term and long-term components. The future level of the time series, at time t, is given by the sum of the contemporaneous effect, the effects of shocks that occurred at times t-1, t-2, t-3, and the effects of shocks on past levels of the series at times t-1, t-2, t-3.
The coefficients of the autoregressive part are within the range {-1, 1}, indicating that each shock tends to be absorbed over time, thus resulting in a temporary level change. The presence of the long-term component allows for the assumption that the effect of the post-COVID recovery lasted beyond the end of the state of emergency, contributing to the increase in spending in 2022.
Regarding the effect of the war in Ukraine, it is modeled only by the short-term component, as its effect diminishes after a lag. In this case, the level of the series at time t is determined solely by the sum of the effects of the shocks that occurred at times t and t-1 on the series level.
Looking at the model diagnostics, the residuals are found to be uncorrelated and seem to follow a random pattern with a symmetric and approximately normal distribution.

Figure10: Plot of fitted values from the model



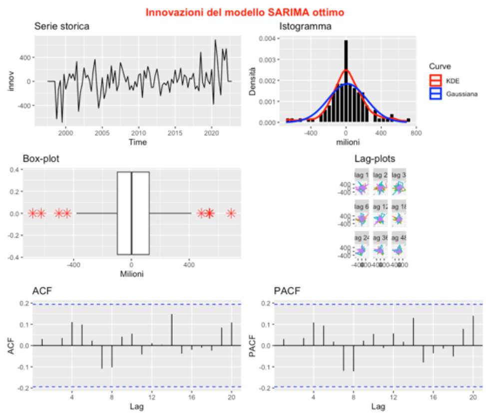
The following graph is obtained by overlaying the estimated values with the fitted values from the optimal model, which explains 90% of the variability.

3.5.2 SARIMA MODEL NORTHEAST SPENDING

Applying the same procedure to the Northeast variable, we observe that the optimal model for modeling the pre-intervention sample is an ARIMA (1,1,2) (1,1,1), and the final model obtained is as follows:

Output3: Post-intervention Model Northeast Spending

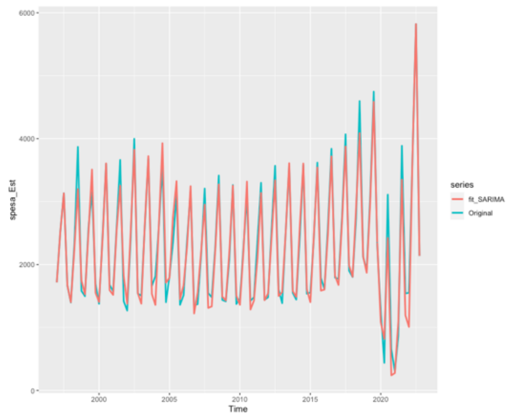
```
Coefficients:
          ar1     ma1      ma2      sar1    sma1   pulse-MA0   pulse-MA1  dummy1-MA0
      -0.6692  0.0705  -0.5001  -0.8830  0.6545  -1032.1738  -1474.8589    931.5580
s.e.   0.3251  0.3124   0.1946   0.1118  0.1697    179.3482    191.7131    302.3806
      dummy1-MA1  dummy1-MA2
      -2419.2071   -1326.062
s.e.    331.9158     268.016

z test of coefficients:

              Estimate  Std. Error z value  Pr(>|z|)
ar1         -6.6918e-01  3.2505e-01 -2.0587 0.0395238 *
ma1          7.0514e-02  3.1238e-01  0.2257 0.8214119
ma2         -5.0005e-01  1.9458e-01 -2.5699 0.0101731 *
sar1        -8.8299e-01  1.1183e-01 -7.8961 2.878e-15 ***
sma1         6.5446e-01  1.6975e-01  3.8555 0.0001155 ***
pulse-MA0   -1.0322e+03  1.7935e+02 -5.7551 8.657e-09 ***
pulse-MA1   -1.4749e+03  1.9171e+02 -7.6931 1.437e-14 ***
dummy1-MA0   9.3156e+02  3.0238e+02  3.0807 0.0020648 **
dummy1-MA1  -2.4192e+03  3.3192e+02 -7.2886 3.132e-13 ***
dummy1-MA2  -1.3261e+03  2.6802e+02 -4.9477 7.510e-07 ***
```



In this case, the ARMA filter applied to the variable modeling the effect of the pandemic consists solely of the moving average component. It is interesting to note that in northeastern Italy, each shock had a short-term impact with a faster absorption rate compared to western Italy. In this geographic area, the effect of COVID seems to have fully dissipated starting from the second quarter of 2022. The sharp increase in prices recorded in 2022 appears to be, in addition to an increase in tourist flows, primarily due to the price hikes resulting from the war in Ukraine.

It is observed that the impact of the war on the increase in spending is significant, and after two lags, this effect diminishes. In conclusion, the residuals seem to be uncorrelated and approximately symmetric.

Figure 11: Plot of fitted values from the model



The resulting model shows a good fit with a $R^2 = 0.85$.

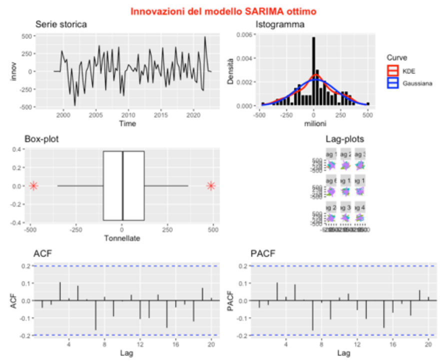
3.5.3 SARIMA MODEL CENTRAL SPENDING

Regarding the spending in central Italy, the optimal model, evaluated on the pre-intervention sample, is an ARIMA (0,1,2) (0,1,2), and the final output with the addition of the new regressors is as follows:

Output4: Post-intervention Model Central Spending

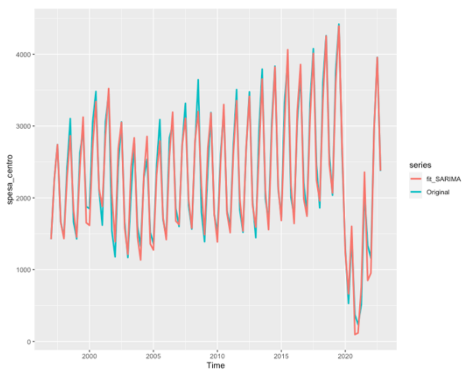
```
Coefficients:
          ma1      ma2     sma1     sma2  pulse-MA0   pulse-MA1  pulse-MA2  pulse-MA3  pulse-MA4  pulse-MA5  pulse-MA6
      -0.3233  -0.4404  -0.4391  -0.3482  -912.3300  -2201.2865   389.3697   267.7117   142.1767  -726.1254  1229.9881
s.e.   0.1091   0.1324   0.1427   0.1350   180.8209    206.1239   156.5965   185.0794   197.6338   230.5341   208.6186
      dummy1-MA0  dummy1-MA1
        -42.0236  -1622.7749
s.e.    269.3555    273.1936
```

```
              Estimate  Std. Error  z value  Pr(>|z|)
ma1           -0.32329     0.10911  -2.9630 0.0030469 **
ma2           -0.44035     0.13243  -3.3252 0.0008836 ***
sma1          -0.43914     0.14271  -3.0771 0.0020900 **
sma2          -0.34817     0.13504  -2.5783 0.0099290 **
pulse-MA0   -912.32996   180.82085  -5.0455 4.524e-07 ***
pulse-MA1  -2201.28646   206.12393 -10.6794 < 2.2e-16 ***
pulse-MA2    389.36970   156.59655   2.4865 0.0129024 *
pulse-MA3    267.71170   185.07941   1.4465 0.1480456
pulse-MA4    142.17672   197.63380   0.7194 0.4718977
pulse-MA5   -726.12542   230.53410  -3.1498 0.0016341 **
pulse-MA6   1229.98809   208.61860   5.8959 3.727e-09 ***
dummy1-MA0   -42.02360   269.35548  -0.1560 0.8760209
dummy1-MA1 -1622.77488   273.19355  -5.9400 2.850e-09 ***
```



As before, the ARMA filter applied to the variable modeling the effect of the pandemic and the war consists solely of the moving average component. The effect of each shock generated by COVID-19 tends to fade after six lags, indicating that the overall impact of COVID is still in the process of being absorbed, as also observed in the graphical analysis.
Analyzing the characteristics of the residuals, they appear to be uncorrelated with an approximately normal distribution.

Figure 12: Plot valori fittati dal modello



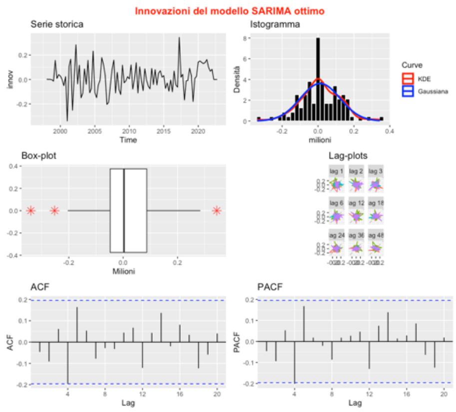
Overall, the model has a good fit with an $R^2 = 0.85$.

3.5.4 SARIMA MODEL SOUTHERN ITALY AND ISLANDS SPENDING

Finally, regarding the logarithm of the spending in southern Italy, the pre-intervention sample is modeled using an ARIMA (1,1,0) (0,1,0). The effect of COVID, modeled using an ARMA (3,4) filter, shows a slower absorption compared to the spending in north-western Italy, as evidenced by the order of the moving average component and the graphical analysis.

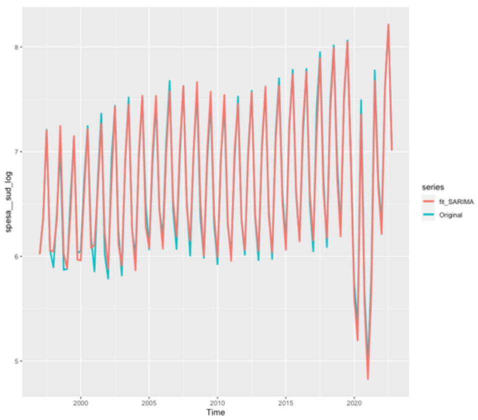
Output5: Post-intervention Model Southern Italy Spending

```
Coefficients:
          ar1  pulse-AR1  pulse-AR2  pulse-AR3  pulse-MA0  pulse-MA1  pulse-MA2  pulse-MA3  pulse-MA4  dummy1-MA0
      -0.2958    -0.5547    -0.5590    -0.8461    -0.7145    -2.1149     0.1006    -1.5822    -1.3972      0.0689
s.e.   0.0993     0.0883     0.0821     0.0762     0.1057     0.1221     0.1646     0.1902     0.1730      0.1777
      dummy1-MA1  dummy1-MA2
         -2.5116      0.9552
s.e.      0.1810      0.1873

           Estimate Std. Error  z value  Pr(>|z|)
ar1        -0.295789   0.099255  -2.9801  0.002881 **
pulse-AR1  -0.554698   0.088343  -6.2789 3.409e-10 ***
pulse-AR2  -0.559005   0.082136  -6.8059 1.004e-11 ***
pulse-AR3  -0.846131   0.076153 -11.1109 < 2.2e-16 ***
pulse-MA0  -0.714494   0.105747  -6.7566 1.412e-11 ***
pulse-MA1  -2.114873   0.122124 -17.3174 < 2.2e-16 ***
pulse-MA2   0.100640   0.164592   0.6114  0.540902
pulse-MA3  -1.582221   0.190183  -8.3195 < 2.2e-16 ***
pulse-MA4  -1.397193   0.172988  -8.0768 6.647e-16 ***
dummy1-MA0  0.068924   0.177736   0.3878  0.698171
dummy1-MA1 -2.511587   0.180958 -13.8794 < 2.2e-16 ***
dummy1-MA2  0.955197   0.187302   5.0998 3.401e-07 ***
```



In conclusion, the distribution of the residuals appears to be approximately symmetric and normal. No significant correlation emerges, except for the fourth lag, which is still negligible.

Figure 13: Plot of fitted values from the model



The model identified above is able to explain 90% of the total variability of the series.

4. REGRESSION

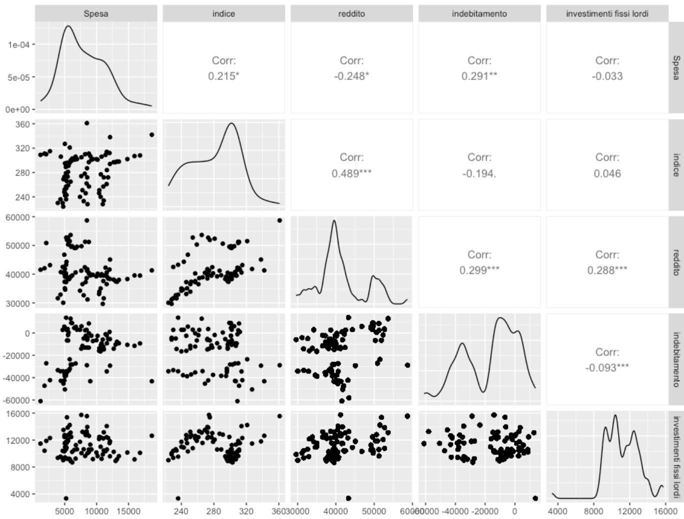
Before the advent of the COVID-19 pandemic, spending by international tourists in Italy experienced a gradual increase. In this chapter, we aim to identify the main economic determinants that may have led to an increase in overall spending.

The covariates analyzed are:

- "Consumer Price Index for the entire population"
- "Indebtedness"
- "Gross fixed investments"
- "Income from dependent work"

Before regressing spending on the covariates under examination, it is necessary to perform the pre-processing phase.

Figure 14: Correlatios

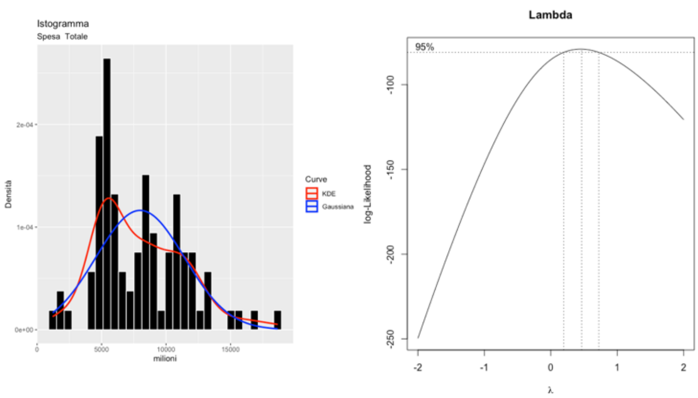


The correlation analysis between the variables allows for the assessment of the presence of collinear variables to be eliminated during the estimation phase, as they generate distorted standard errors.

The graph highlights the absence of collinear variables and the lack of non-linearity between the covariates and the spending in Italy.

Next, we proceed by studying the distribution of the dependent variable to assess appropriate transformations aimed at improving its normality, linear relationship, and heteroscedasticity.

Figure 15: histogram of total spending and Box-Cox plot



The total spending by tourists in Italy shows a positive skew, with an empirical density function that significantly deviates from the theoretical normal distribution. By applying the Box-Cox method to the raw estimated model, the dependent variable is transformed using a square root transformation.

To quantify the effect of the SARS COVID-19 pandemic on the spending by international tourists, a dichotomous dummy variable is added to the model, which takes the value of 0 when the event is absent and 1 when the event is present:

$$\text{Dummy} = \begin{cases} 1\ if\ 2020\ I \le\ t \le 2022\ I \\ 0\ otherwise \end{cases}$$

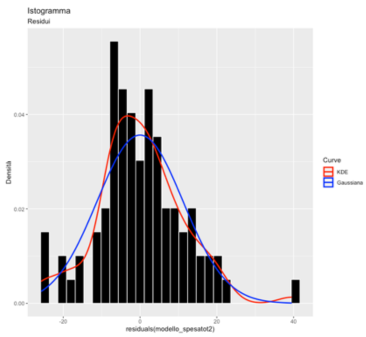
The estimated model is as follows:

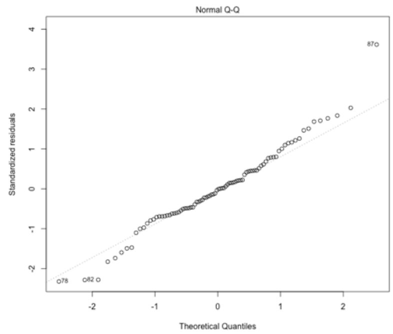
Output6: regression linear model

```
Residuals:
    Min      1Q  Median      3Q     Max
-19.318  -7.667  -0.834   5.832  32.456

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                 3.723e+01  1.356e+01   2.745  0.00742 **
indice                      5.389e-01  5.044e-02  10.683  < 2e-16 ***
reddito                    -2.931e-03  2.736e-04 -10.712  < 2e-16 ***
indebitamento               7.763e-04  9.016e-05   8.610 4.30e-13 ***
`investimenti fissi lordi`  2.947e-03  6.773e-04   4.351 3.88e-05 ***
covid                      -1.565e+01  4.715e+00  -3.319  0.00135 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.82 on 82 degrees of freedom
Multiple R-squared:  0.7051,    Adjusted R-squared:  0.6871
F-statistic:  39.2 on 5 and 82 DF,  p-value: < 2.2e-16
```

The output shows that the range of variability of the model's residuals is quite high, and the median value does not perfectly coincide with the mean, indicating the presence of positive skewness. The Q-Q plot and the histogram confirm the results obtained from the variability range, i.e., the non-normal distribution of the residuals. From this preliminary analysis, it can be concluded that the estimated model is not robust. To make it robust, a more in-depth analysis of the concept of linearity would be needed, identifying optimal transformations of the covariates through non-parametric methods to maximize the linear relationship between the independent variables and the target variable. Another crucial procedure is the detection of outliers, i.e., anomalous values that cause the non-normality of the residual distribution and the presence of heteroscedasticity.

Although the model is not robust, a preliminary interpretation can be made regarding the relationship between the independent variables and the dependent variable. All the covariates are statistically significant, indicating the existence of a causal relationship with spending. The latter shows a positive relationship with all the covariates except for income. Of particular interest is the relationship between spending and the inflation index, a factor that, net of other variables, has had the most significant impact on the increase in tourist spending.





To interpret the beta coefficients of the continuous variables in relation to the target variable expressed in its original scale, it is necessary to calculate the marginal effects as follows:

$$\frac{dy}{dx} = 2(\beta_j^2)x_j$$

Regarding the impact of COVID, it is possible to calculate the percentage change in spending using the following formula:

$$\frac{Y_{covid} - Y_{nocovid}}{Y_{nocovid}} = \frac{(a^2 + b^2 + 2ab) - (a^2)}{a^2} = -0.66$$

Thus, the pandemic seems to have negatively impacted the average spending level, reducing it by approximately 66%.

5. CONCLUSION

In conclusion, the impact of the pandemic on tourism spending trends differs across geographical areas. The region most affected by the preventive measures introduced starting from the first quarter of 2022 is central Italy, which is still in the process of economic recovery. On the other hand, the region that managed to absorb the pandemic-related damage the fastest was the eastern area.

Starting from the second quarter of 2022, despite the number of tourists being lower than in 2019, spending increased significantly in all areas of Italy, except for the center. This sudden growth can be attributed not only to the post-pandemic economic recovery but also to the impact of the outbreak of the war in Ukraine, which caused an increase in fuel and electricity prices, leading to higher inflation.

This work, therefore, serves as a preliminary study to outline the basic and main characteristics of the phenomenon under investigation. It would be advisable to deepen the analysis by regressing the spending trends not only based on its past levels but also on additional covariates that can better explain the fluctuations and the temporal progression of the series.