

*ANALISI DELLA VARIAZIONE
DELLA SPESA SOSTENUTA DAI
TURISTI INTERNAZIONALI IN
ITALIA A SEGUITO DELLA
PANDEMIA DA COVID 19*

Sara Capozio
Matricola 853600

1. INTRODUZIONE

Il settore turistico è stato tra i settori maggiormente colpiti dalla pandemia COVID-19. La limitazione degli spostamenti e la chiusura delle attività turistiche hanno determinato una flessione significativa dei flussi turistici nel territorio italiano.

L'obiettivo dell'analisi consiste nel valutare se l'impatto della pandemia sulla spesa sostenuta dai turisti internazionali è risultato differente tra le seguenti aree territoriali: Nord-Est, Nord-Ovest, Centro, Sud e Isole e se tale effetto si è riassorbito negli anni successivi all'emergenza globale.

Il dataset in esame comprende tre serie storiche che coprono un periodo temporale di 26 anni con cadenza trimestrale. Le rilevazioni partono dal primo trimestre del 1997 fino a concludersi il quarto trimestre del 2022. La variabile oggetto dell'analisi è la ‘Spesa per area territoriale’ che include tutte le spese finalizzate al viaggio, sostenute prima della partenza, durante il viaggio o al ritorno, in conseguenza dello stesso. Questa è composta dalle spese per il trasporto, per l'alloggio, per i pasti, acquisti vari, attività ricreative, culturali e sportive. L'ordine di grandezza di tale variabile è espresso in milioni di euro.

Nella prima parte dell'analisi vengono introdotte due variabili di controllo:

- ‘Viaggiatori a destinazione’: ovvero il conteggio dei viaggiatori nei singoli luoghi visitati. L'ordine di grandezza di tale variabile è espresso in migliaia.
- ‘Pernottamenti per area territoriale’: numero delle notti trascorse dai clienti negli esercizi ricettivi nel periodo considerato. L'ordine di grandezza di tale variabile è espresso in migliaia.

Queste due variabili sono utilizzate a fine interpretativo per comprendere al meglio il comportamento della spesa nell'arco temporale d'interesse.

Un secondo obiettivo dell'analisi che verrà sviluppato in modo approfondito nel capitolo della regressione lineare consiste nell'identificazione dei principali determinanti che hanno generato un possibile aumento o diminuzione della spesa complessiva sostenuta dai turisti in Italia.

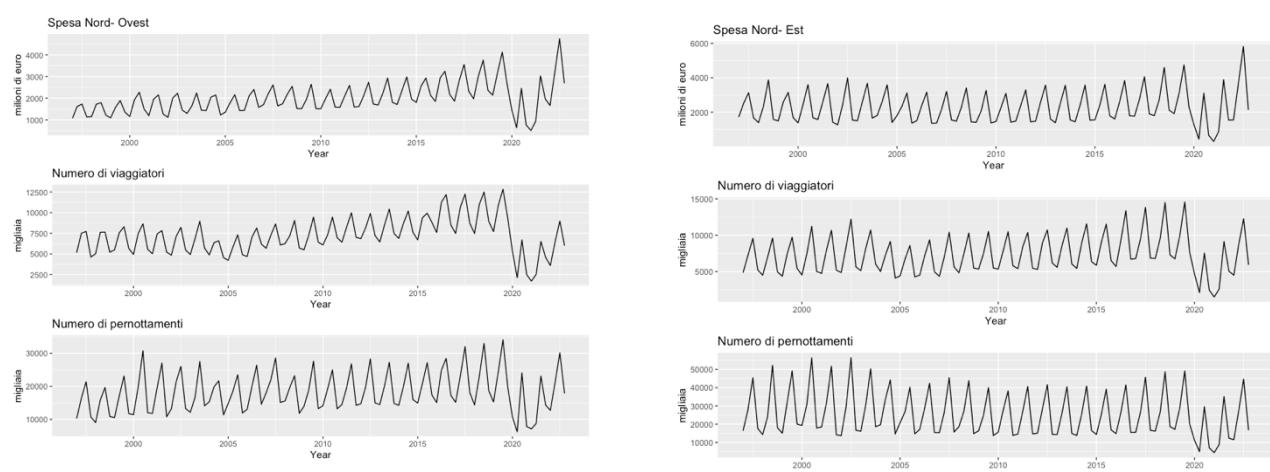
Le covariate analizzate di natura economica sono:

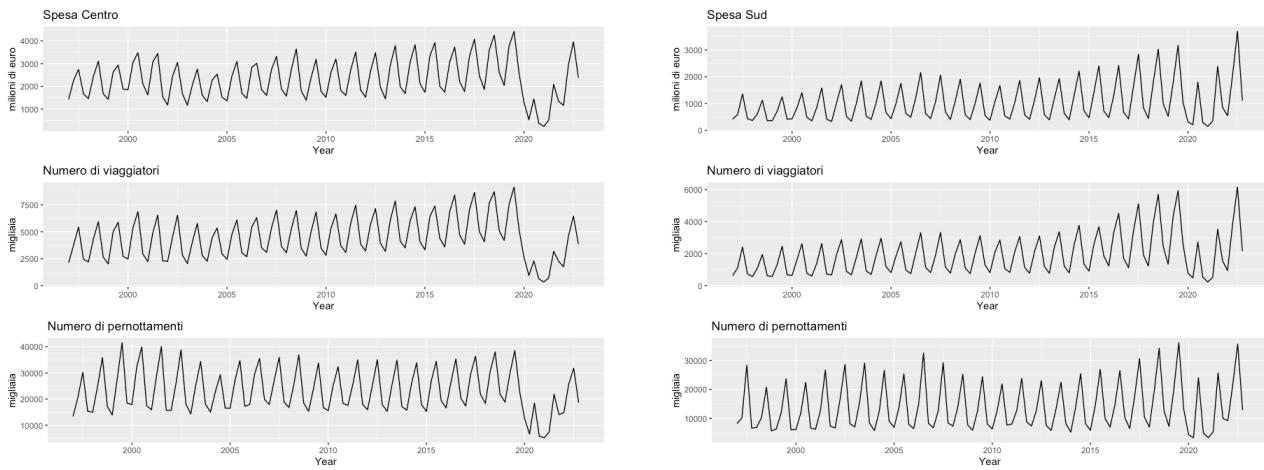
- "indice dei prezzi al consumo per l'intera collettività": misura statistica ottenuta attraverso la media dei prezzi di un insieme di beni e servizi, utilizzata generalmente come indicatore d'inflazione. Tale variabile è espressa in milioni di euro
- “indebitamento”: Saldo contabile tra le entrate e le uscite. Tale variabile è espressa in milioni di euro
- “investimenti fissi lordi”: acquisti di beni materiali durevoli effettuati da un'impresa nell'esercizio che comprendono l'acquisto di macchine, impianti, attrezzature, mobili, mezzi di trasporto, costruzioni e fabbricati, terreni e l'incremento di capitali fissi per lavori interni. L'ordine di grandezza di tale variabile è espresso in milioni di euro.
- "reddito da lavoro dipendente": retribuzioni percepite in dipendenza di un rapporto di lavoro subordinato. L'ordine di grandezza di tale variabile è espresso in milioni di euro.

2. ANALISI GRAFICA DELLE SERIE STORICHE

Analizzando graficamente le tre serie storiche in esame per ogni macroregione risulta possibile identificare a priori eventuali componenti non osservabili, pattern particolari o eventuali break strutturali.

Figura1: rappresentazione grafica serie storiche





Dalla Figura 1 possiamo osservare la presenza di una forte componente stagionale e di una variazione repentina temporanea (break strutturale) avvenuta nel primo trimestre dell'anno 2020 causata dall'epidemia di Covid 19. Per quanto riguarda l'andamento temporale delle variabili, le serie storiche che caratterizzano la spesa e i viaggiatori sono differenti per area territoriale.

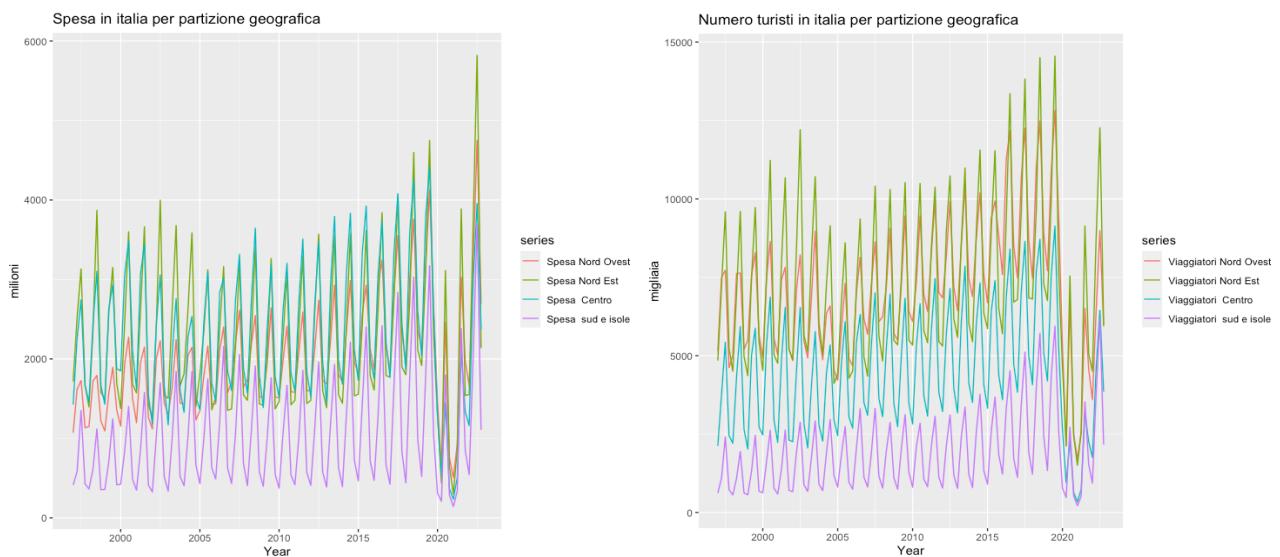
Nella zona Nord-Ovest e del Centro Italia la spesa sostenuta dai turisti esteri presenta un andamento lineare crescente, mentre nell'area Nord Est sembrerebbe seguire un lieve andamento parabolico. Per quanto riguarda il Sud Italia, il fenomeno appare essere eteroschedastico con variabilità direttamente proporzionale alla dimensione temporale; tuttavia, i dati non seguono una particolare forma funzionale.

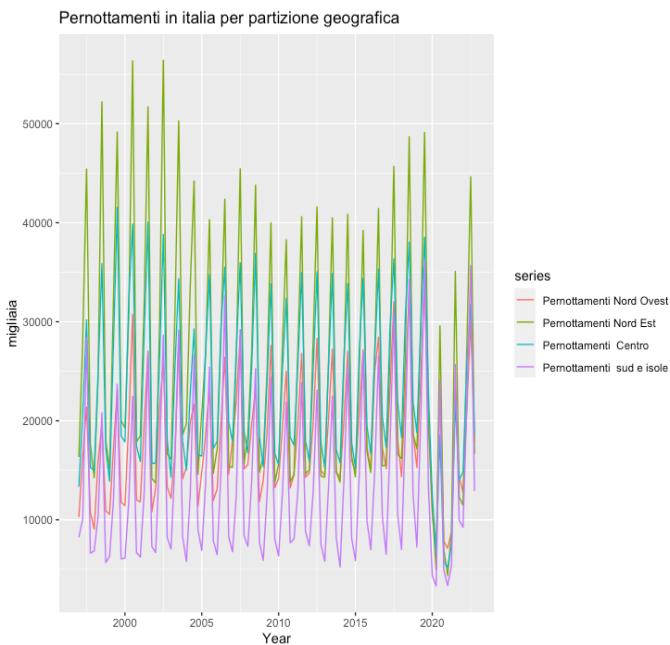
La serie storica relativa al numero di viaggiatori, sia nell'area settentrionale che centrale del Paese, presenta un importante cambio di pendenza registrato nel 2005 che ha comportato un costante aumento nel numero di turisti all'interno delle aree geografiche di riferimento, mentre nell'Italia meridionale si è rilevato un costante aumento del numero di turisti solamente a partire dal 2015.

Nonostante il numero di viaggiatori sia aumentato nel corso del tempo, il numero di pernottamenti in tutte le zone eccetto l'area Nord-Ovest è rimasto costante non evidenziando significativi incrementi o decrementi.

Per individuare analogie o differenze tra le aree geografiche e valutarne l'impatto dell'epidemia si sovrappongono le serie storiche.

Figura 2: rappresentazione grafica serie storiche sovrapposte





Nel 2020, con l'avvento della pandemia da SARS Covid-19 la spesa complessiva dei viaggiatori stranieri in Italia è diminuita complessivamente di 17,3 miliardi di euro. Dai grafici possiamo notare che le aree maggiormente colpite dalla pandemia sono le aree settentrionali e centrali. Si osserva che l'effetto del covid sul turismo è stato riassorbito in tutte le zone, fatta eccezione per il centro Italia che risulta essere la zona più colpita. Il settore turistico in quest'ultima area è ancora in fase di ripresa presentando, infatti, nel quarto trimestre dell'anno 2022 un numero di turisti e una spesa complessiva inferiore rispetto al quarto trimestre dell'anno 2019. Anche il mezzogiorno ha ripreso in positivo l'attività turistica rilevando nel 2022 un aumento del numero di visitatori e di pernottamenti.

3. PROCEDURA DI BOX-JENKINS

Tramite la metodologia di Box & Jenkins, risulta possibile identificare il modello ottimale in grado di approssimare al meglio l'andamento temporale della spesa per area geografica.

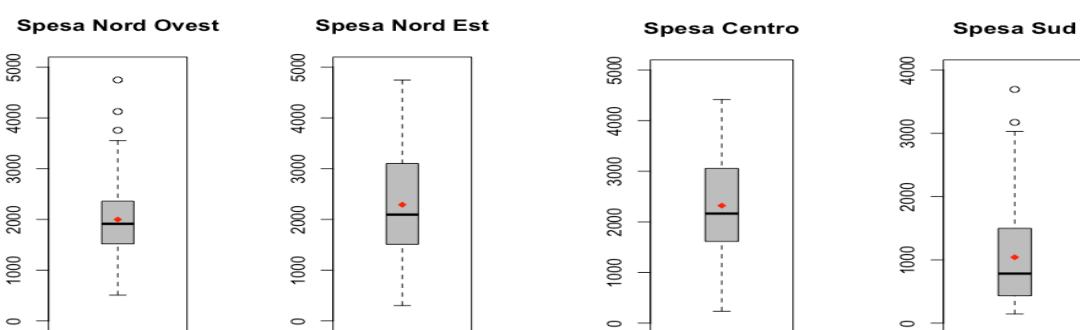
La procedura si articola nelle seguenti fasi:

1. Analisi esplorativa
2. Trasformazione dei dati
3. Analisi della stazionarietà
4. Analisi delle componenti della serie
5. Stima del modello e valutazione delle performance di adattamento

3.1 ANALISI ESPLORATIVA

L'analisi esplorativa permette di identificare il tipo di distribuzione delle variabili in esame e l'eventuale esistenza di punti outliers.

Figura3: box-plot



Considerando la finestra temporale prima dell'avvento della pandemia, le regioni del Nord Est e Ovest intercettano la gran parte dei flussi turistici internazionali. Anche la spesa dei turisti italiani è notevolmente concentrata a livello geografico, con il Nord Est e il centro Italia che assorbono più della metà della spesa totale.

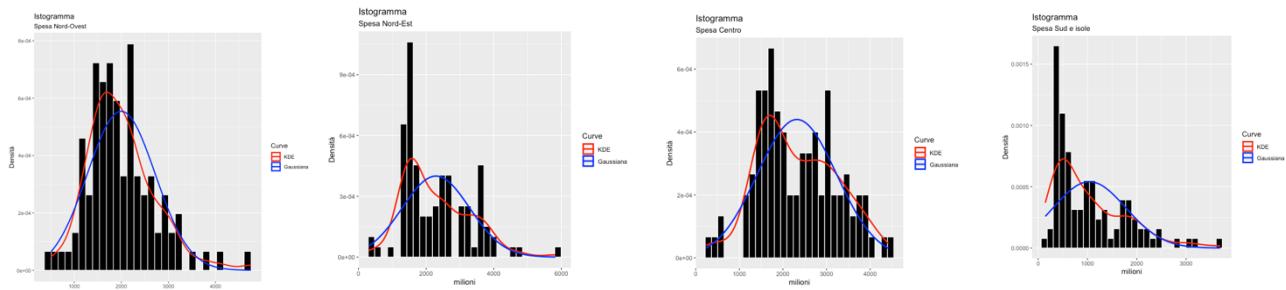
Dai grafici possiamo notare come l'area geografica centrale, nonostante registri un numero di visitatori inferiore, rispetto a quello della zona nord Occidentale, presenti una spesa media complessiva nettamente superiore pari a 322.9 milioni di euro, rappresentando una delle mete più costose per i turisti esteri.

Inoltre, risulta evidente la discordanza fra i flussi di viaggiatori internazionali e il potenziale turistico che caratterizza il Mezzogiorno. Quest'ultima rappresenta la regione con il numero più basso di visitatori assorbendo solo il 10% della spesa totale.

Analizzando i box-plot, le variabili in esame sono tutte caratterizzate da un'asimmetria positiva in quanto la media risulta essere significativamente maggiore della mediana. Si individua inoltre, la presenza di valori anomali che si discostano dalla distribuzione della spesa relativa alle macroregioni Nord-Ovest e Sud. Analizzando nel dettaglio i valori outliers, essi si riferiscono alla forte crescita della spesa sostenuta dai turisti nell'area Nord Occidentale nel terzo trimestre del 2019 e del 2022. Si registra, infatti, un aumento rispettivamente del 10% e del 30% rispetto al 2018. Allo stesso modo, nel Mezzogiorno, il terzo trimestre del 2019 e del 2022 ha presentato un aumento della spesa pari al 5% e 20% rispetto al 2018.

L'analisi degli istogrammi raffigurati conferma l'ipotesi di asimmetria positiva dei dati, in quanto le distribuzioni di densità empiriche per ogni variabile di ogni macroregione si discostano dalla forma funzionale normale teorica.

Figura4: istogrammi



Le evidenze ottenute tramite l'analisi grafica sembrerebbero concordare con i risultati generati dai test di normalità di Bera-Jarque e Shapiro Wilk.

Tabella 1: Tabella test Bera-Jarque

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

Osservando la Tabella1, si può notare che, per ogni livello di significatività, si ha sufficiente evidenza empirica per rifiutare l'ipotesi nulla che le distribuzioni abbiano un indice di simmetria nullo ed un valore dell'indice di curtosi pari a tre; solamente la spesa nel Centro Italia sembrerebbe presentare una distribuzione simmetrica.

Tabella 2: Tabella test Shapiro-Wilk

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

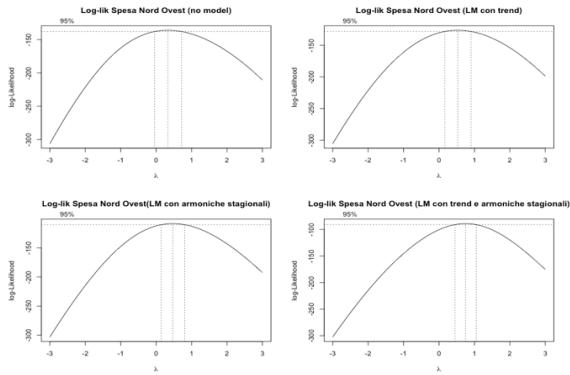
Il test di Shapiro Wilk, basato sulla correlazione tra i dati empirici e i valori teorici di normalità, si rifiuta l'ipotesi nulla di normalità dei dati per ogni livello di significatività.

Nonostante la 'Spesa' relativa al centro Italia sia simmetrica, la sua distribuzione non segue l'andamento di una distribuzione normale. Tale risultato si può dedurre osservando lo scostamento delle funzioni di densità teorica ed empirica rappresentate dagli istogrammi.

3.2 TRASFORMAZIONE DEI DATI

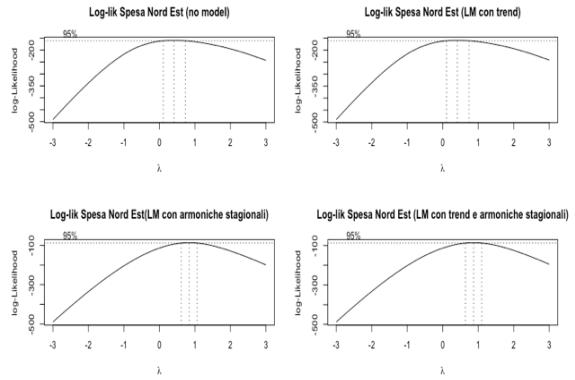
Uno strumento utile che permette di risolvere il problema di non normalità dei dati e l'eventuale presenza di eteroschedasticità è la trasformazione di Box-Cox. L'obiettivo di tale trasformazione consiste nell'identificare una funzione stabilizzatrice della varianza in modo tale che le osservazioni che compongono la serie storica abbiano variabilità costante al variare della dimensione temporale.

Figura 5: Grafici Box-Cox

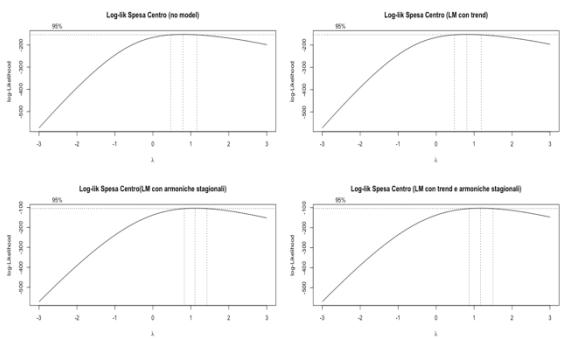


La figura 5 rappresenta i valori di lambda ottimali per la spesa relativa alla zona nord-occidentale ottenuti applicando il metodo di Guerrero e il metodo di massimizzazione della log-verosimiglianza profilo di diversi modelli in cui la variabile target è regredita in funzione del trend, della stagionalità e della loro combinazione. Osservando il grafico relativo alla spesa complessiva sostenuta dai turisti, tutti e quattro gli intervalli di confidenza costruiti ad un livello di significatività del 95% contengono il valore $\frac{1}{2}$ suggerendo di applicare la trasformazione in radice quadrata alla variabile d'interesse.

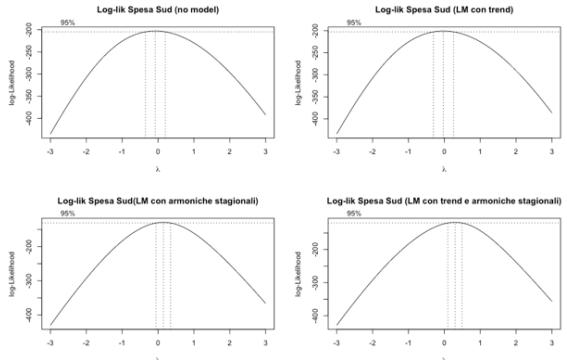
Per quanto riguarda invece la spesa nell'area nordorientale, gli intervalli di confidenza sembrerebbero essere molto variabili. Si decide quindi di focalizzarsi sul quarto modello costruito su dati destagionalizzati e detrendizzati in quanto, depurando le variabili dalle relative componenti, esso ne cattura l'andamento sottostante. Poiché l'intervallo di confidenza contiene il valore unitario, si opta per non applicare alcun tipo di trasformazione.



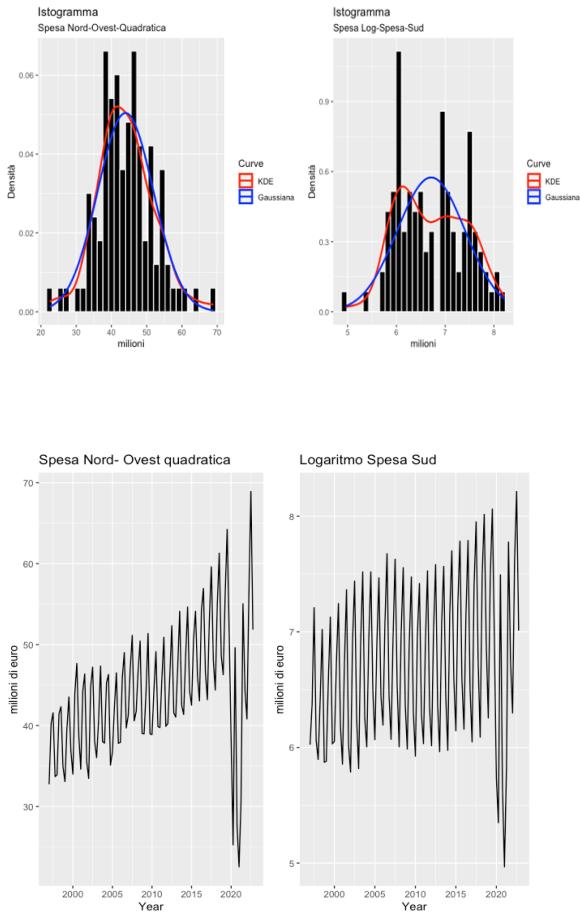
Analizzando la spesa nell'Italia centrale, gli intervalli di confidenza costruiti attorno al valore di lambda ottimale sono coerenti tra loro e, poiché contengono il valore unitario, la variabile non richiede alcun tipo di trasformazione.



Infine, considerando l'area Meridionale possiamo osservare che gli intervalli di confidenza costruiti attorno al parametro lambda ottimale relativi alla variabile 'Spesa' contengono lo zero, suggerendo di apportare una trasformazione logaritmica ai dati.



Una volta individuata le opportune trasformazioni da effettuare analizziamo le distribuzioni e i grafici delle variabili trasformate:



Osservando l'istogramma, la trasformazione applicata alla variabile ‘Spesa Nord-Ovest’ sembrerebbe aver migliorato l’asimmetria della sua distribuzione rendendola normale.

Il risultato ottenuto è confermato dal valore del p-value, pari a 0.2314, applicando il test di Bera-Jarque.

Per quanto riguarda la trasformazione logaritmica applicata alla variabile spesa meridionale, la distribuzione della variabile trasformata sembrerebbe essere simmetrica, anche se la sua funzione di densità empirica non coincide perfettamente con quella teorica di una normale. Il test di Bera-Jarque conferma i risultati riscontrati dall’analisi grafica ottenendo un p-value pari a 0.1749.

Le trasformazioni in radice quadrata e logaritmica, analizzando l’andamento temporale delle variabili trasformate, sembrerebbero aver stabilizzato le fluttuazioni stagionali delle serie prima del break avvenuto nel 2020.

3.3 ANALISI DELLA PERSISTENZA

Per analizzare la stazionarietà delle serie storiche si effettua il test di Dickey Fuller in modo da verificare l’ipotesi nulla di presenza di una radice unitaria, ovvero l’esistenza di un trend stocastico. Supponiamo che la serie storica X_t sia interpretabile con un processo $AR(p)$ che si muove attorno ad una componente deterministica CD_t . La verifica d’ipotesi assume diverse configurazioni in base alla forma della componente deterministica:

- $CD_t = 0$, ovvero la componente deterministica è nulla
- $CD_t = k$, ovvero la componente deterministica è una costante diversa da zero
- $CD_t = k + \delta t$, ovvero la componente deterministica è lineare rispetto al tempo

Verranno quindi effettuati tre test differenti, il primo per testare se il processo è un processo stazionario con media nulla, il secondo per valutare la stazionarietà intorno ad una costante e, infine, il terzo per valutare la stazionarietà attorno ad un trend deterministico.

La tabella 3 mostra i risultati ottenuti applicando il test di Dickey Fuller ipotizzando nulla la componente deterministica.

Il test verifica il seguente sistema d’ipotesi:

$$\left\{ \begin{array}{l} H_0: \text{il processo è un random walk senza drift} \\ H_1: \text{il processo è un processo stazionario} \\ \quad \quad \quad \text{con media nulla} \end{array} \right.$$

Tabella 3: Tabella Test 1 Dickey-Fuller

Test 1 Dickey-Fuller			
	1 pct	5 pct	10 pct
SPESA OVEST: Statistica test = -0.653	tau1= -2.58	tau1= -1.95	tau1= -1.62
SPESA EST: Statistica test = -2.3456	1 pct tau1= -2.58	5 pct tau1= -1.95	10 pct tau1= -1.62
SPESA CENTRO: Statistica test = -2.2096	1 pct tau1= -2.58	5 pct tau1= -1.95	10 pct tau1= -1.62
SPESA SUD: Statistica test = -0.5567	1 pct tau1= -2.58	5 pct tau1= -1.95	10 pct tau1= -1.62

Osservando i risultati in tabella, i valori delle statistiche test per l'area nord-occidentale e meridionale ricadono nella regione di accettazione dell'ipotesi nulla per ogni livello di significatività. Per quanto riguarda invece il Centro e il Nord Est i valori della statistica test sono inferiori al valore critico corrispondente al livello di significatività del 5% e 10%, ma superiori a quello dell'1%.

La tabella 4 mostra i risultati ottenuti applicando il test di Dickey Fuller ipotizzando la presenza di un drift, ovvero una componente deterministica diversa da zero. Il test verifica il seguente sistema d'ipotesi:

$$\left\{ \begin{array}{l} H_0: \text{il processo è un random walk} \\ H_1: \text{il processo è un processo stazionario} \\ \quad \text{con media diversa da zero} \end{array} \right.$$

Tabella 4: Tabella Test2 Dickey-Fuller

Test 2 Dickey-Fuller				
	Tau2: -6.996	1 pct -3.46	5 pct -2.88	10 pct -2.57
NORD OVEST	Phi1: 24.4981	1 pct 6.52	5 pct 4.63	10 pct 3.81
NORD EST	Tau2: -12.7435	1 pct -3.46	5 pct -2.88	10 pct -2.57
CENTRO	Phi1: 81.2004	1 pct 6.52	5 pct 4.63	10 pct 3.81
SUD	Tau2: -10.3593	1 pct -3.46	5 pct -2.88	10 pct -2.57
	Phi1: 53.6575	1 pct 6.52	5 pct 4.63	10 pct 3.81
	Tau2: -14.718	1 pct -3.46	5 pct -2.88	10 pct -2.57
	Phi1: 108.3201	1 pct 6.52	5 pct 4.63	10 pct 3.81

Osservando i valori in tabella si nota la presenza di due statistiche test. La statistica test tau2 verifica che nel modello autoregressivo, includendo una costante diversa dal valore nullo, vi sia una radice unitaria. La statistica test phi1 verifica l'ipotesi nulla congiunta che il processo sia non stazionario e che la costante sia pari a zero.

Poiché in tutte e quattro le variabili i valori delle statistiche test cadono nelle regioni di rifiuto per ogni livello di significatività, si conclude che il processo è stazionario attorno ad un drift.

Infine, si verifica il caso in cui la componente deterministica è un trend lineare e dunque il seguente sistema d'ipotesi:

$$\left\{ \begin{array}{l} H_0: \text{il processo è un random walk con drift} \\ H_1: \text{il processo è un processo stazionario attorno ad un trend} \end{array} \right.$$

Tabella 5: Tabella Test 3 Dickey-Fuller

Test 3 Dickey-Fuller				
		1 pct	5 pct	10 pct
NORD OVEST	Tau3: -8.5844 2	-3.99	-3.43	-3.13
	Phi2: 24.5914	6.22	4.75	4.07
	Phi3: 36.8571	8.43	6.49	5.47
	Tau3: -12.7519	-3.99	-3.43	-3.13
	Phi2: 54.2128	6.22	4.75	4.07
	Phi3: 81.3172	8.43	6.49	5.47
NORD EST	Tau3: -10.3467	-3.99	-3.43	-3.13
	Phi2: 35.6847	6.22	4.75	4.07
	Phi3: 53.5269	8.43	6.49	5.47
	Tau3: -16.0537	-3.99	-3.43	-3.13
	Phi2: 85.9153	6.22	4.75	4.07
	Phi3: 128.8608	8.43	6.49	5.47
CENTRO	Tau3: -10.3467	-3.99	-3.43	-3.13
	Phi2: 35.6847	6.22	4.75	4.07
	Phi3: 53.5269	8.43	6.49	5.47
	Tau3: -16.0537	-3.99	-3.43	-3.13
	Phi2: 85.9153	6.22	4.75	4.07
	Phi3: 128.8608	8.43	6.49	5.47
SUD	Tau3: -10.3467	-3.99	-3.43	-3.13
	Phi2: 35.6847	6.22	4.75	4.07
	Phi3: 53.5269	8.43	6.49	5.47
	Tau3: -16.0537	-3.99	-3.43	-3.13
	Phi2: 85.9153	6.22	4.75	4.07
	Phi3: 128.8608	8.43	6.49	5.47

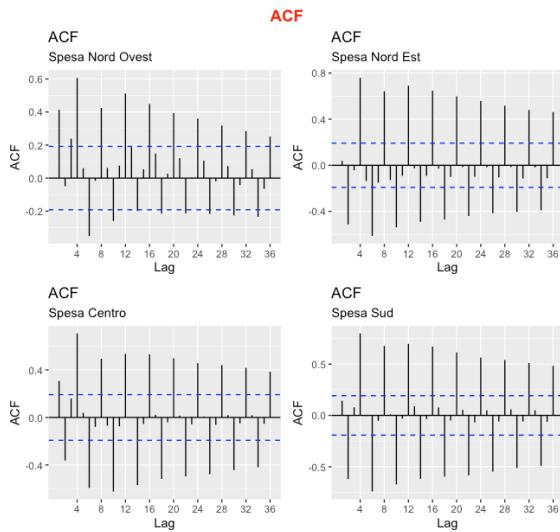
Osservando i valori in tabella si nota la presenza di tre statistiche test.

La statistica test tau3 verifica che nel modello autoregressivo, che include un trend lineare, ci sia una radice unitaria, la statistica test phi3 verifica l'ipotesi nulla congiunta che il processo sia non stazionario e che la slope del trend sia nulla. Infine, la statistica test phi2 verifica che il processo sia non stazionario e che i parametri dell'equazione del trend siano congiuntamente nulli.

Anche in questo caso i valori delle statistiche test cadono nelle regioni di rifiuto per ogni livello di significatività e si può concludere che le variabili sono stazionarie attorno ad un trend deterministico, evidenza già riscontrata tramite l'analisi grafica.

Terminata l'analisi della stazionarietà, si procede ad esaminare la persistenza delle serie storiche tramite l'ausilio del correlogramma (ACF) e del correlogramma parziale (PACF) che mostrano la variazione delle correlazioni tra le osservazioni al variare dei ritardi.

Figura6: Grafico ACF

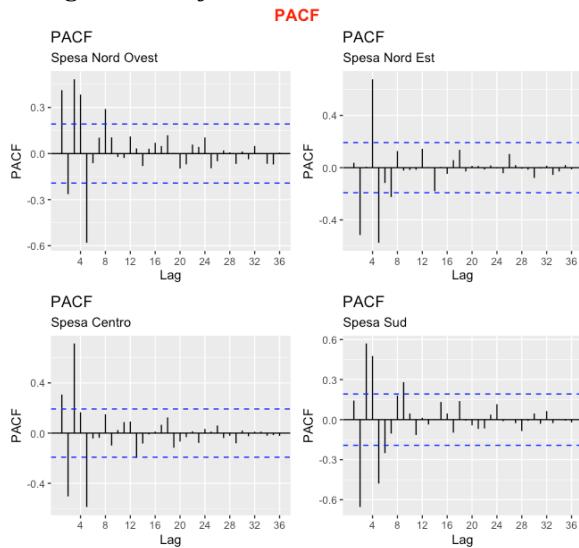


Analizzando la figura 6 si osserva che le correlazioni tra i lags sono significativamente diverse da zero con un livello di significatività del 95%, in quanto fuoriescono dalle bande di confidenza definite dalle linee tratteggiate blu.

In tutti i grafici la correlazione misurata al ritardo quattro è maggiore rispetto agli altri ritardi. Questo è dovuto alla componente stagionale presente nei dati; i picchi, infatti, tendono a ripetersi ogni quattro lags.

Si nota, inoltre, che all'aumentare dei ritardi l'autocorrelazione tende a diminuire, questo è dovuto all'effetto della componente trend presente nei dati.

Figura7: Grafico PACF



Per quanto riguarda il grafico delle autocorrelazioni parziali, depurate dagli effetti intermedi, si osserva che fino al ritardo 5 le autocorrelazioni parziali sono significativamente diverse da zero per poi dal ritardo 6 riassorbirsi completamente.

3.4 ANALISI DELLE COMPONENTI DELLE SERIE STORICHE

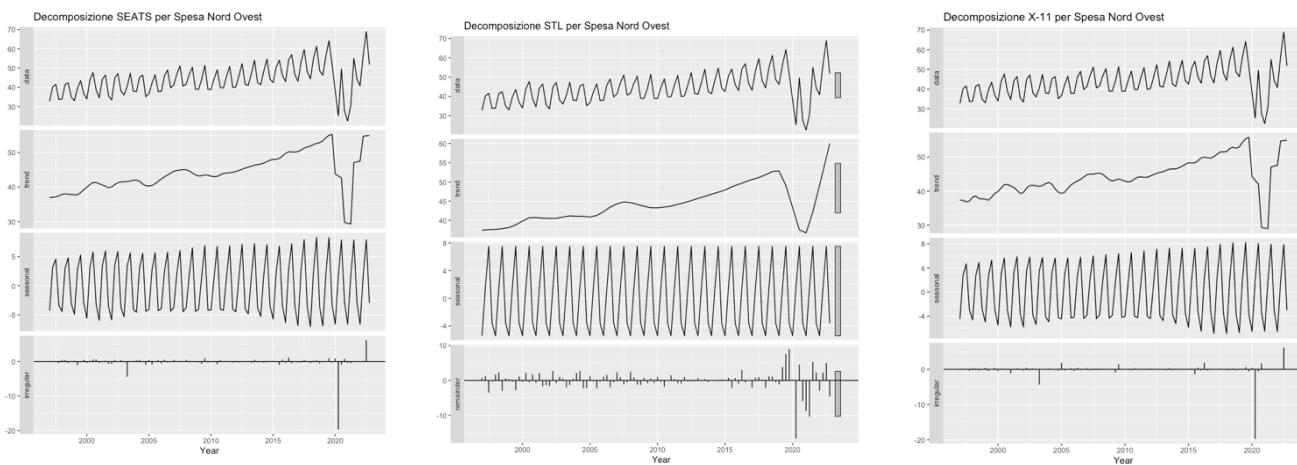
Dalle analisi effettuate, l'andamento delle variabili nel tempo sembrerebbe essere composto dall'aggregazione di tre componenti non osservabili:

- Trend
- Stagionalità
- Break strutturale

Le variabili in esame non sono dei fenomeni eteroschedastici in quanto, la magnitudine delle fluttuazioni stagionali, non varia all'aumentare dell'ampiezza temporale, affermazione confermata dai test di Dickey-Fuller. Dunque, la relazione esistente tra le serie storiche e le singole componenti sembrerebbe essere di tipo additivo.

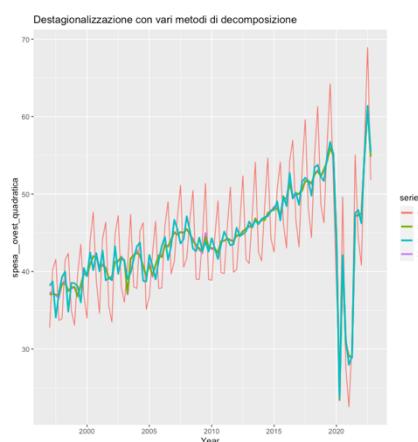
La decomposizione delle serie è stata effettuata tramite metodi robusti alla presenza di outliers e cambi di livello, ossia attraverso X-11, TRAMO SEATS e STL.

Figura8: Grafico decomposizioni

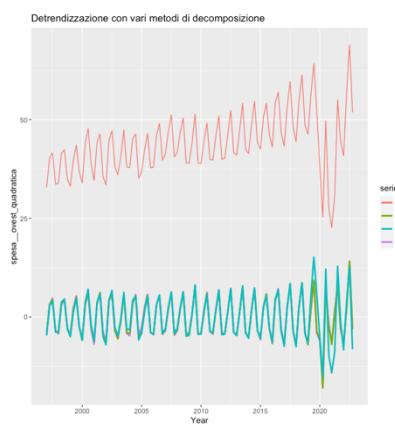


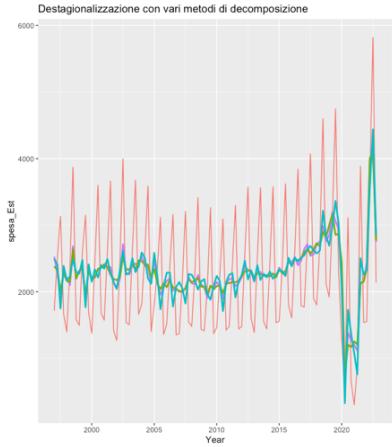
La figura 8 evidenzia che tutti e tre metodi di decomposizione catturano la forte componente stagionale, il trend e la presenza del break strutturale. Il metodo STL, a differenza degli altri due algoritmi, consente di catturare al meglio i valori outliers; questo grazie alla sua capacità di monitorare le componenti ciclo-trend e stagionalità tramite un parametro di smoothing. I metodi illustrati, depurando la serie dal trend e dalla stagionalità, non sono in grado di carpire completamente il break strutturale che diventa così una componente residua.

Figura9: Grafico serie destagionalizzate e detrendizzate



La sovrapposizione delle serie destagionalizzate relative alla ‘Spesa’ nord-occidentale, riportate in figura 9, mostra la similarità dei metodi di decomposizione. Maggiori discrepanze si osservano analizzando le serie detrendizzate. Infatti, come precedentemente citato, solo il metodo STL è riuscito a catturare al meglio l’andamento degli shock presenti dal 2020 in poi.

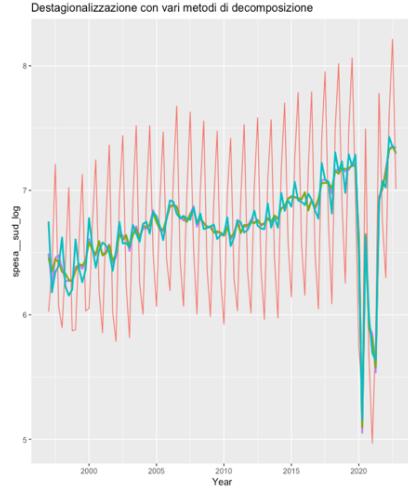




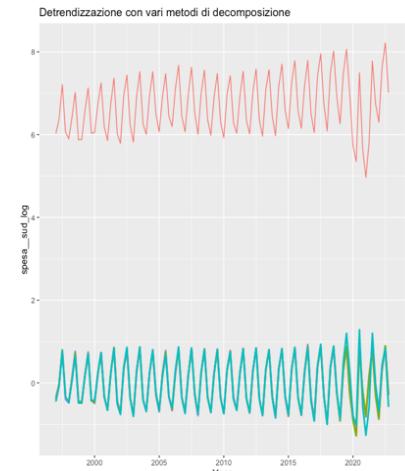
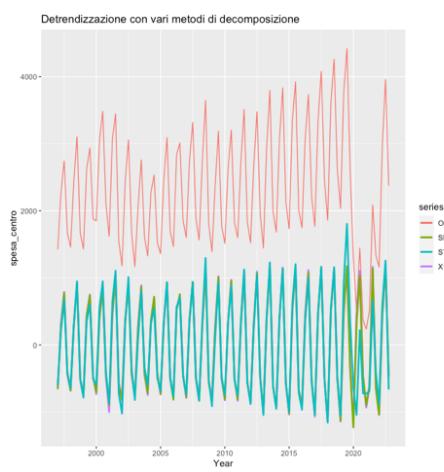
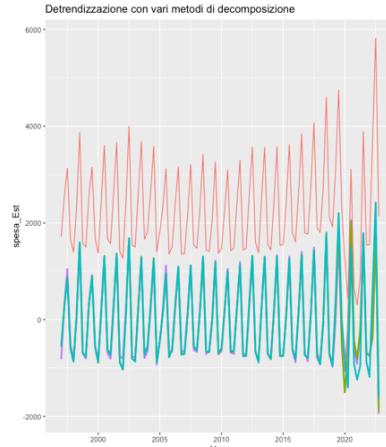
Le considerazioni effettuate sulla ‘Spesa’ nord-occidentale sono valide anche per le altre variabili in esame. In generale, il processo di destagionalizzazione basato sul parametro di smoothing del metodo STL, rende le serie destagionalizzate meno lisce rispetto agli altri metodi, in quanto la velocità di variazione della componente stagionale può essere controllata dall’utente stesso.



Per quanto riguarda il processo di detrendizzazione, tutti e tre i metodi, depurando le serie dal trend, perdono gran parte delle informazioni contenute nella componente break strutturale. Il metodo che riesce al meglio a catturare gli shock risulta essere il metodo STL.

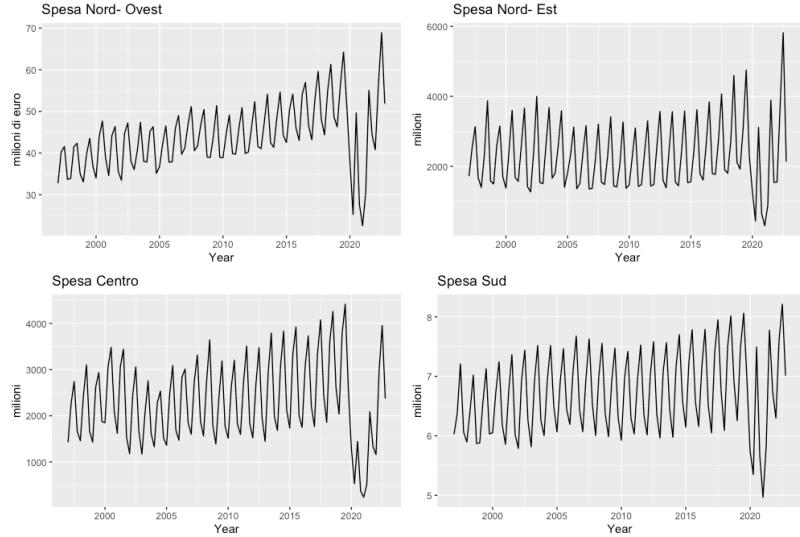


Nel caso della ‘Spesa’ nell’italia del mezzogiorno e isole i metodi di decomposizione sembrerebbero equivalersi.



3.5 MODELLI SARIMA

Analizzando l'andamento temporale della spesa sostenuta dai turisti in Italia, si osserva la presenza di due significativi break strutturali. Il primo break coincide con l'avvento della pandemia da COVID 19, pandemia diffusasi a livello globale a partire dal dicembre 2019. Lo shock generato dalla prima ondata ha avuto un risultato evidente sulla variazione della spesa del trimestre successivo generando un repentino cambio di livello della serie storica. L'effetto dell'emergenza globale ha avuto un impatto ancora più drammatico sul turismo invernale dell'anno 2021, questo dovuto all'introduzione di nuove misure restrittive in seguito ad un aumento del numero di casi registratosi nell'estate 2021. Con la dichiarazione della fine dello stato di emergenza, il 31 marzo del 2022, l'effetto della pandemia in Italia Nord orientale, occidentale e meridionale, sembrerebbe essersi riassorbito raggiungendo approssimativamente lo stesso livello di spesa registratosi nel 2019. Situazione opposta si è verificata nel centro Italia, area maggiormente colpita dagli effetti della pandemia e ancora in fase di ripresa economica, dove nel primo trimestre del 2022 si è registrato un livello della spesa inferiore del 40% rispetto al 2019. A partire dal secondo trimestre del 2022, una forte crescita della spesa ha coinvolto, con più o meno intensità, tutte le aree territoriali nonostante quest'ultime negli ultimi anni abbiano presentato un flusso turistico inferiore rispetto al 2019. Questo repentino aumento oltre ad essere causato dalla ripresa economica in seguito alla pandemia, può essere dovuto dall'impatto che ha avuto lo scoppio della guerra in Ucraina nel febbraio del 2022 sull'aumento dei prezzi delle materie prime tra cui energia e carburante. Nell'Italia orientale, occidentale e meridionale la spesa ha raggiunto il suo livello massimo aumentando rispettivamente del 15%, 22%, 16% rispetto al valore registrato il terzo trimestre dell'anno 2019. Questa crescita ha riguardato, anche se non in modo significativo, l'Italia centrale registrando comunque una spesa inferiore del 10% rispetto a quella rilevata nei tre anni precedenti.



Per estrarre i due shock si è effettuata un intervention analysis, ovvero una procedura modellistica per incorporare gli effetti di eventi o interventi esogeni che hanno agito sulla serie.

Nella prima fase, la tecnica automatizzata di Hyndman & Khandakar identifica il modello *SARIMA* ottimale per il campione pre-intervento (pre-covid) per poi applicarlo, nella seconda fase, su tutte le osservazioni di cui è composta la serie. In quest'ultimo step al modello vengono aggiunti due nuovi regressori:

- Una prima variabile dummy dicotomica per identificare il periodo dello stato d'emergenza

$$X_{1t} = \begin{cases} 1 & \text{se } 2020 I \text{ (inizio pandemia)} \leq t \leq 2022 I \text{ (dichiarazione fine stato emergenza)} \\ 0 & \text{altrimenti} \end{cases}$$

- Una seconda variabile dummy per modellare l'effetto dello scoppio della guerra in Ucraina nel febbraio del 2022 sull'aumento della spesa complessiva

$$X_{2t} = \begin{cases} 1 & \text{se } 2022 II \text{ (inizio guerra)} \leq t \leq 2022 IV \text{ (fine della guerra)} \\ 0 & \text{altrimenti} \end{cases}$$

A quest'ultime, è stato applicato un filtro *ARMA* in modo da modellare l'impatto degli shock sul livello della spesa e la loro velocità di riassorbimento nel breve e lungo periodo.

Alla prima variabili dummy si è assegnato valore unitario per l'intero periodo della pandemia e non solo in corrispondenza del suo scoppio in quanto, dal febbraio del 2020 fino a marzo del 2022, gli enti governativi mondiali sono intervenuti con numerose norme preventive che hanno impattato in modo differente su livello

della spesa turistica. Medesimo ragionamento è stato effettuato per la variabile dummy associata alla guerra, in quanto si vuole identificare, non solo l'effetto diretto dello shock contemporaneo dovuto allo scoppio dello scontro, ma anche tutti gli effetti che ha prodotto nel tempo come i rincari energetici, l'incremento del prezzo del petrolio e l'inflazione.

3.5.1 MODELLO SARIMA SPESA NORD OVEST

Applicando la procedura sopra esposta alla spesa registrata in Italia nord-occidentale si ottiene il seguente output:

Output1: modello pre-intervento Spesa Nord Ovest

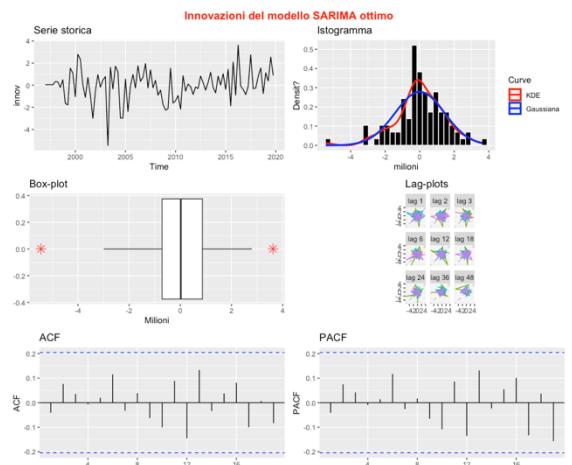
z test of coefficients:

```
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



Il modello ottimale per modellare il campione pre-intervento è un modello *ARIMA* (1,1,0) (0,1,1). Dalle diagnostiche del modello si nota che l'andamento dei residui è molto prossimo ad un andamento casuale; inoltre, non vi è alcun tipo di autocorrelazione tra i lags. Per quanto riguarda la loro distribuzione, essi sembrerebbero distribuirsi approssimativamente come una normale anche se la presenza di un punto outlier sulla coda di sinistra rende la distribuzione leggermente asimmetrica negativa.

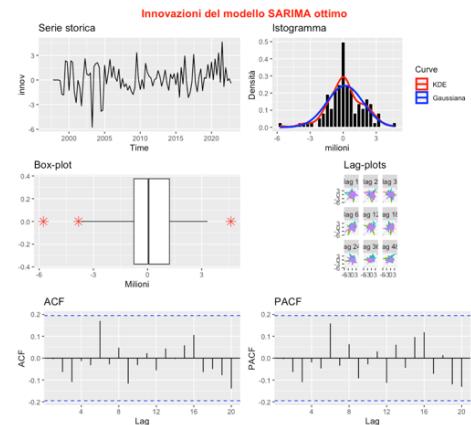
Output2: modello post-intervento Spesa Nord Ovest

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
```



Il modello dinamico individuato è il seguente:

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

$$\text{dove } X_{1t} = \begin{cases} 1 & \text{se } 2020 I \leq t \leq 2022 I \\ 0 & \text{altrimenti} \end{cases} \quad X_{2t} = \begin{cases} 1 & \text{se } 2022 II \leq t \leq 2022 IV \\ 0 & \text{altrimenti} \end{cases}$$

Si osserva che l'effetto della pandemia è modellata dalla componente a breve e lungo periodo.

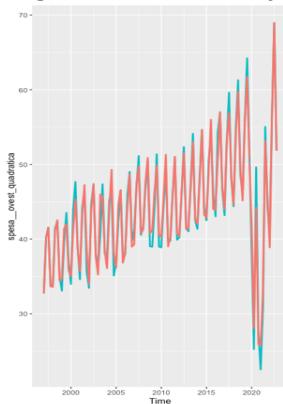
Il livello futuro della serie storica, al tempo t , è dato dalla somma dell'effetto contemporaneo, degli effetti degli shock avvenuti al tempo $t - 1, t - 2, t - 3$ e degli effetti degli shock sui livelli passati della serie storica al tempo $t - 1, t - 2, t - 3$.

I coefficienti della parte autoregressiva sono compresi nel range $\{-1, 1\}$ stando a identificare che ogni shock tende a riassorbirsi nel tempo comportando così un cambio di livello temporaneo. La presenza della componente di lungo periodo permette di supporre che l'effetto della ripresa post-covid si è prolungato anche oltre la fine dello stato di emergenza contribuendo all'aumento della spesa del 2022.

Per quanto riguarda l'effetto della guerra in Ucraina, esso è modellato solo dalla componente di breve periodo in quanto dopo un lag l'effetto decade. In questo caso il livello della serie al tempo t è dato solamente dalla somma degli effetti degli shock avvenuti al tempo t e $t - 1$ sul livello della serie.

Osservando invece le diagnostiche del modello, i residui risultano essere incorrelati e sembrerebbero seguire un andamento casuale con distribuzione simmetrica e approssimativamente normale.

Figura10: Plot valori fittati dal modello

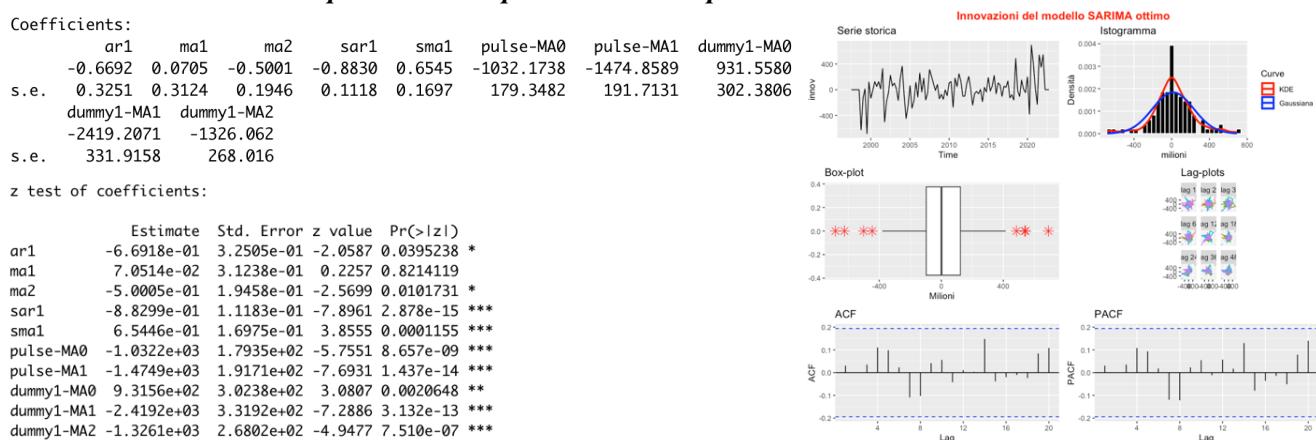


Il seguente grafico è ottenuto sovrapponendo i valori stimati con i valori fittati dal modello ottimale con una variabilità spiegata del 90%

3.5.2 MODELLO SARIMA SPESA NORD EST

Applicando la medesima procedura alla variabile Nord-Est osserviamo che il modello ottimale che modella il campione pre-intervento è un *ARIMA* (1,1,2) (1,1,1) e il modello finale ottenuto è il seguente:

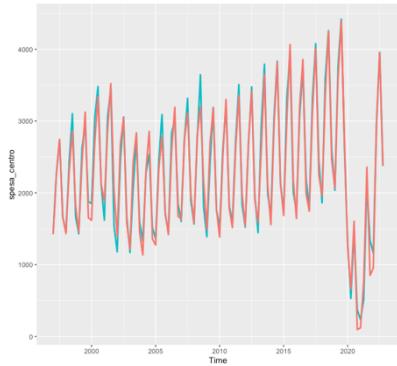
Output3: modello post-intervento Spesa Nord Est



In questo caso il filtro *ARMA* applicato alla variabile che modella l'effetto della pandemia è composto solamente dalla componente a media mobile. Risulta interessante osservare che in Italia nord-orientale ogni shock ha avuto un impatto a breve termine con una velocità di riassorbimento maggiore rispetto all'Italia occidentale. In questa area geografica l'effetto del Covid pare essersi riassorbito completamente a partire dal secondo trimestre del 2022. Il forte aumento dei prezzi registrato nel 2022 sembrerebbe essere dovuto, oltre ad un aumento dei flussi turistici, principalmente al rincaro prezzi conseguente alla guerra in Ucraina.

Si nota, infatti, che l'impatto della guerra sull'aumento della spesa è significativo e dopo due lags tale effetto decade. In conclusione, i residui sembrerebbero essere incorrelati e approssimativamente simmetrici.

Figura11: Plot valori fittati dal modello



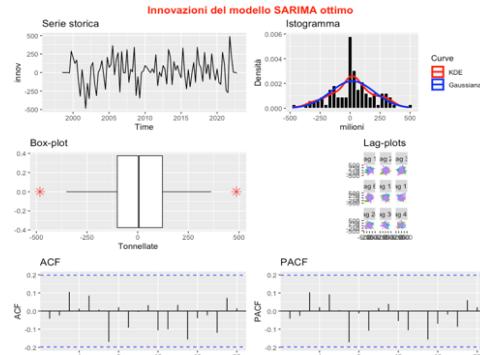
Il modello risultante registra un buon adattamento con un $R^2 = 0.85$.

3.5.3 MODELLO SARIMA SPESA CENTRO

Per quanto riguarda la spesa sostenuta nel centro Italia il modello ottimale, valutato sul campione pre-intervento, è un *ARIMA* (0,1,2) (0,1,2) e l'output finale con l'aggiunta dei nuovi regressori risulta essere:

Output4: modello post-intervento Spesa centro

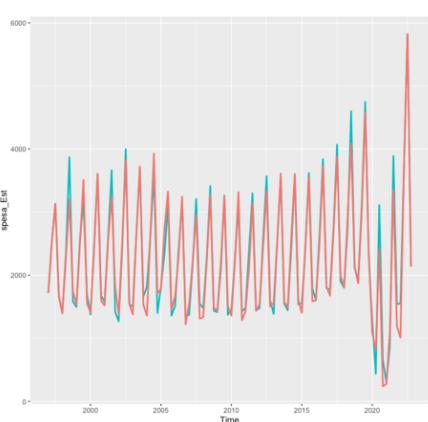
Coefficients:										
ma1	ma2	sma1	sma2	pulse-MA0	pulse-MA1	pulse-MA2	pulse-MA3	pulse-MA4	pulse-MA5	pulse-MA6
-0.3233	-0.4404	-0.4391	-0.3482	-912.3300	-2201.2865	389.3697	267.7117	142.1767	-726.1254	1229.9881
s.e.	0.1091	0.1324	0.1427	0.1350	180.8209	206.1239	156.5965	185.0794	197.6338	230.5341
dummy1-MA0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



Come in precedenza, il filtro ARMA applicato alla variabile che modella l'effetto della pandemia e della guerra è composto solamente dalla componente a media mobile. L'effetto di ogni shock generato dal COVID 19 tende a decadere dopo sei lags stando ad indicare che l'impatto complessivo del covid è ancora in fase di riassorbimento, evidenza riscontrata anche dall'analisi grafica.

Analizzando le caratteristiche dei residui, sembrerebbero essere incorrelati con una distribuzione approssimativamente normale.

Figura12: Plot valori fittati dal modello



Complessivamente il modello ha un buon adattamento con un $R^2 = 0.85$.

3.5.4 MODELLO SARIMA SPESA SUD E ISOLE

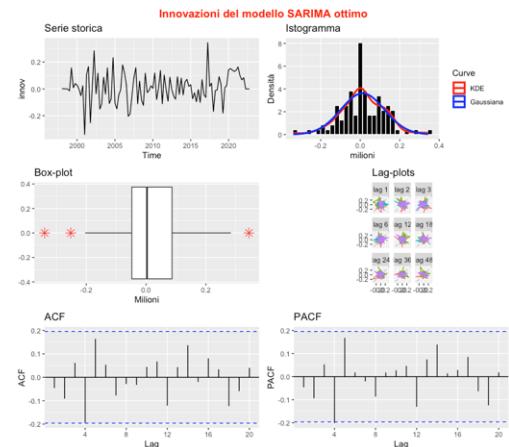
Infine, per quanto riguarda il logaritmo della spesa sostenuta nel sud-Italia, il campione pre-intervento è modellato tramite un *ARIMA* (1,1,0) (0,1,0). L'effetto del covid, modellato tramite un filtro *ARMA* (3,4), presenta un riassorbimento più lento rispetto alla spesa nord-occidentale, evidenza osservabile dall'ordine della componente a media mobile e dall'analisi grafica.

Output5: modello post-intervento Spesa Sud

Coefficients:

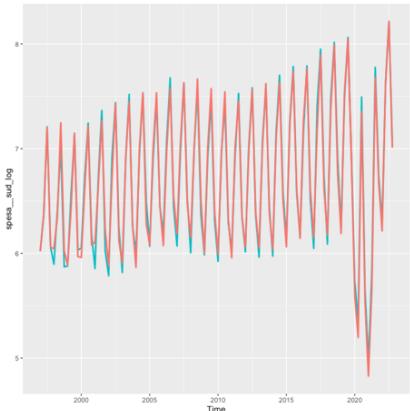
	ar1	pulse-AR1	pulse-AR2	pulse-AR3	pulse-MA0	pulse-MA1	pulse-MA2	pulse-MA3	pulse-MA4	dummy1-MA0
s.e.	0.0993	0.0883	0.0821	0.0762	0.1057	0.1221	0.1646	0.1902	0.1730	0.1777
dummy1-MA1		dummy1-MA2								
	-2.5116	0.9552								
s.e.	0.1810	0.1873								

	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 ***



Concludendo, la distribuzione dei residui pare essere approssimativamente simmetrica e normale. Non emerge nessun tipo di correlazione significativa ad eccezione del quarto ritardo che risulta essere comunque trascurabile.

Figura13: Plot valori fittati dal modello



Il modello sopra identificato è in grado di spiegare il 90% della variabilità totale della serie.

4. REGRESSIONE

Precedentemente l'avvento della pandemia COVID-19, la spesa sostenuta dai turisti internazionali in Italia ha subito una progressiva crescita.

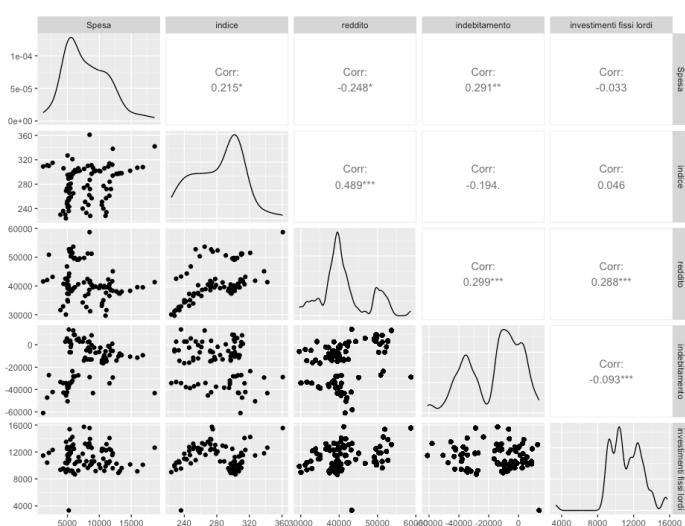
In questo capitolo, si vogliono identificare i principali determinanti di natura economica che hanno generato un possibile aumento della spesa complessiva.

Le covariate analizzate sono:

- “indice di prezzi al consumo per l'intera collettività”
- “indebitamento”
- “investimenti fissi lordi”
- “reddito da lavoro dipendente”

Prima di regredire la spesa in funzione delle covariate in esame è necessario svolgere la fase di pre-processing.

Figura14: Grafico correlazioni

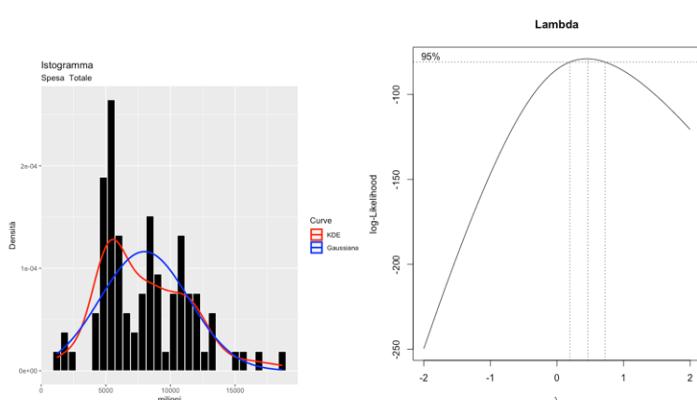


L'analisi delle correlazioni tra le variabili permette di valutare la presenza di variabili collineari da eliminare in fase di stima in quanto generano standard error distorti.

Il grafico evidenzia l'assenza di variabili collineari e la non linearità tra le covariate e la spesa sostenuta in Italia.

Si procede studiando la distribuzione della variabile dipendente per valutare opportune trasformazioni al fine di migliorarne la normalità, la relazione di linearità e l'eteroschedasticità.

Figura15: istogramma spesa totale e grafico Box-Cox



La spesa complessiva sostenuta dai turisti in Italia presenta un'asimmetria positiva con una funzione di densità empirica che si discosta significativamente da quella normale teorica. Applicando il metodo di Box-Cox al modello grezzo stimato, si procede trasformando la variabile dipendente tramite trasformazione in radice quadrata.

Per quantificare l'effetto della pandemia da SARS COVID-19 sulla spesa sostenuta dai turisti internazionali si inserisce all'interno del modello una variabile dummy dicotomica che assuma valore 0 quando l'evento è assente e 1 quando l'evento è presente:

$$\text{Dummy} = \begin{cases} 1 & \text{se } 2020 \leq t \leq 2022 \\ 0 & \text{altrimenti} \end{cases}$$

Il modello stimato è il seguente:

Output6: modello regressione lineare

L'output mostra come il campo di variazione dei residui del modello sia molto elevato e il valore della mediana non coincide perfettamente con il valore della media, rilevando così la presenza di una forma di asimmetria positiva. Il grafico q-q plot e l'istogramma confermano i risultati ottenuti dal range di variabilità, ovvero la non normalità distributiva dei residui. Da questa analisi preliminare si può concludere che il modello stimato non sia un modello robusto. Per renderlo tale, bisognerebbe effettuare un'analisi più approfondita sul concetto di linearità identificando trasformazioni ottimali delle covariate, tramite metodi non parametrici, al fine di massimizzare il legame lineare tra le variabili indipendenti e la variabile target.

Altra procedura fondamentale è la rilevazione di punti outliers, ovvero valori anomali che determinano la non normalità della distribuzione dei residui e la presenza di eteroschedasticità. Nonostante il modello non sia robusto, è possibile dare un'interpretazione preliminare sulla relazione tra le variabili indipendenti e la variabile dipendente.

Tutte le covariate sono statisticamente significative stando ad indicare l'esistenza di un legame di causalità con la spesa. Quest'ultima presenta un legame positivo con tutte le covariate ad eccezione del reddito. D'interesse è la relazione tra la spesa e l'indice d'inflazione, fattore che, al netto delle altre variabili, ha impattato maggiormente sull'aumento della spesa turistica.

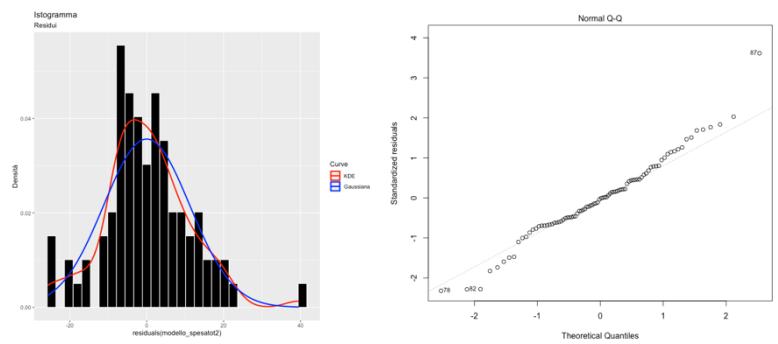
Per interpretare i coefficienti beta delle variabili continue, in relazione alla variabile target espressa in scala originale, è necessario calcolare gli effetti marginali nel seguente modo:

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

Per quanto riguarda l'impatto del COVID, è possibile calcolarne la variazione percentuale sulla spesa tramite la seguente formula:

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

Dunque, la pandemia sembrerebbe aver impattato negativamente sul livello medio spesa riducendola circa del 66%.



5. CONCLUSIONE

In conclusione, l'impatto della pandemia sull'andamento della spesa turistica differisce per aree territoriali. La zona che ha risentito, con maggior intensità, delle norme di prevenzione introdotte a partire dal primo trimestre del 2022 è il centro Italia, attualmente ancora in fase di ripresa economica.

Al contrario, l'area che è riuscita a riassorbire più velocemente i danni generati dalla pandemia è stata quella orientale.

A partire dal secondo trimestre del 2022, nonostante il numero di turisti fosse inferiore rispetto a quello registrato nel 2019, la spesa sostenuta è aumentata significativamente in tutte le zone d'Italia, fatta eccezione per il centro. Questa crescita repentina può essere legata, oltre alla ripresa economica post-pandemica, all'impatto dello scoppio della Guerra in Ucraina che ha causato un rincaro dei prezzi del gasolio e dell'elettricità, determinando conseguentemente un aumento dell'inflazione.

Questo elaborato funge, quindi, da studio preliminare al fine di delineare le caratteristiche base e principali del fenomeno in studio. Sarebbe opportuno approfondire l'analisi regredendo l'andamento della spesa non solo in funzione dei suoi livelli passati ma anche di covariate aggiuntive in grado di spiegare al meglio le fluttuazioni e l'andamento temporale della serie.