

Analisi di Mercato e Stima d’Impatto: Studio sulla Competizione nel Settore Assicurativo Italiano

Valerio Lapiello, Analista Indipendente

19 Luglio 2025

Contents

1	Executive Summary	4
2	Contesto di Business e Dati	4
2.1	Obiettivo dello Studio	4
2.2	Il Contesto Competitivo	4
2.3	L’Evento di Intervento	5
2.4	Fonte Dati e Metrica di Analisi	5
3	Analisi Esplorativa dei Dati (EDA)	5
3.1	Visualizzazione Iniziale e Dinamiche di Trend	5
3.2	Decomposizione e Pattern Stagionali	6
4	Metodologia	7
4.1	Approccio Generale: Analisi di Intervento con Modelli ARIMAX	7
4.2	Pre-processing e Test di Stazionarietà	7
4.3	Identificazione e Selezione dei Modelli	8
5	Risultati e Analisi	8
5.1	Validazione del Pre-processing e Stazionarietà delle Serie	8
5.2	Selezione del Modello per Allianz	9
5.3	Selezione del Modello per Prima Assicurazioni	10
5.4	Stima dell’Impatto della Campagna	11
5.4.1	Auto-Impatto: Un Rallentamento della Crescita per Prima Assicurazioni	11
5.4.2	Impatto Competitivo: Resilienza del Brand Allianz	12
5.5	Analisi Previsionale Comparativa	12

6	Interpretazione di Business e Raccomandazioni Strategiche	12
6.1	Il Vento a Favore: Sfruttare il Momentum del Mercato	13
6.2	Efficacia ed Efficienza della Campagna: Un Insight Contro-Intuitivo	13
6.3	Lo Scenario Futuro: Il Divario si Restringe	14
7	Limiti e Sviluppi Futuri	14
7.1	Limiti Metodologici	14
7.2	Prospettive di Sviluppo Futuro	14
	Appendice A: Dettagli Metodologici e Diagnostica	15
A.1	Grafici di Autocorrelazione (ACF/PACF) delle serie stazionarie	15
A.2	Output di Stima e Diagnostica dei Modelli Candidati per Allianz	16
A.2.1	Modello A Scartato Allianz: SARIMA(0,1,0)(0,1,0)[12]	16
A.2.2	Modello B Vincitore Allianz: SARIMA(1,1,1)(1,0,0)[12]	17
A.3	Output di Stima e Diagnostica dei Modelli Candidati per Prima Assicurazioni	18
A.3.1	Modello A Vincitore Prima: ARIMA(4,1,0) with Drift	18
A.3.2	Modello B scartato Prima: ARIMA(3,1,0)	19
A.4	Output dei Modelli di Intervento Finali	21
A.4.1	Modello ARIMAX Allianz	21
A.4.2	Modello ARIMAX Prima	21
	Appendice B: Codice R Completo e Commentato	22

List of Figures

1	Indice di Interesse Storico Comparativo, Gen 2020 - Giu 2025.	5
2	Decomposizione delle Serie Storiche per Brand. Allianz (sinistra), Prima Assicurazioni (destra)	6
3	Serie Storiche Trasformate e Stazionarie per Brand.	9
4	Analisi dei Residui del modello finale per Allianz	10
5	Analisi dei Residui del modello finale per Prima	11
6	Previsioni a 12 Mesi per Brand con intervalli di Confidenza al 95%.	13
7	A.1 Autocorrelogrammi Allianz stazionaria, indicazione di un possibile Seasonal Random Walk	15
8	A.1 Autocorrelogrammi Prima logaritmica stazionaria, si teorizzava un ARIMA(3,1,0)	15
9	A.2.1 Analisi dei Residui del modello scartato per Allianz	16
10	A.2.1 QQ-Plot dei residui del modello scartato per Allianz.	17
11	A.2.2 QQ-Plot dei residui del modello finale per Allianz.	18
12	A.3.1 QQ-Plot dei residui del modello finale per Prima.	19
13	A.3.2 Analisi dei Residui del modello scartato per Prima	20
14	A.3.2 QQ-Plot dei residui del modello scartato per Prima.	20

List of Tables

1	Statistiche Descrittive Principali per Brand.	6
2	Risultati del Test ADF prima e dopo il Pre-processing.	8
3	Confronto dei Modelli Candidati per Allianz.	9
4	Coefficienti Stimati per il Modello Finale di Allianz (SARIMA(1,1,1)(1,0,0)[12]).	9
5	Confronto dei Modelli Candidati per Prima Assicurazioni (su serie logaritmica).	10
6	Coefficienti Stimati per il Modello Finale di Prima (ARIMA(4,1,0) with drift su log).	11
7	Risultati del Modello di Intervento per Prima Assicurazioni.	11
8	Risultati del Modello di Intervento per Allianz.	12
9	A.2.1 Statistiche di bontà del fit per il modello scartato di Allianz.	16
10	A.2.2 Coefficienti stimati, errori standard e statistiche per il modello finale di Allianz.	17
11	A.2.2 Statistiche di bontà del fit per il modello finale di Allianz.	17
12	A.3.1 Coefficienti stimati, errori standard e statistiche per il modello finale di Prima.	18
13	A.3.1 Statistiche di bontà del fit per il modello finale di Prima.	18
14	A.3.2 Coefficienti stimati, errori standard e statistiche per il modello scartato di Prima.	19
15	A.3.2 Statistiche di bontà del fit per il modello scartato di Prima.	19
16	A.4.1 Coefficienti stimati, errori standard e statistiche per il modello ARIMAX di Allianz.	21
17	A.4.1 Statistiche di bontà del fit per il modello ARIMAX di Allianz.	21
18	A.4.2 Coefficienti stimati, errori standard e statistiche per il modello ARIMAX di Prima.	21
19	A.4.2 Statistiche di bontà del fit per il modello ARIMAX di Prima.	21

1 Executive Summary

Questo studio analizza l'impatto della campagna marketing di Prima Assicurazioni (Maggio 2024) sull'interesse di ricerca online per il proprio brand e per il leader di mercato Allianz, utilizzando modelli di intervento ARIMAX su dati Google Trends.

Emerge un quadro a due velocità. Da un lato, l'analisi previsionale conferma un **forte momentum di crescita per Prima Assicurazioni**, tale da poter ridurre significativamente il divario con l'incumbent nel medio termine. Dall'altro, l'analisi di intervento rivela un risultato contro-intuitivo: la campagna, pur costosa, si è dimostrata **inefficace nell'erodere la quota di interesse di Allianz** e **inefficiente**, avendo rallentato la traiettoria di crescita organica di Prima stessa.

Lo studio evidenzia quindi un mercato in fermento, dove la posizione del leader, sebbene resiliente, è minacciata da una crescita di fondo dei challenger digitali, i quali, a loro volta, necessitano di ottimizzare i propri investimenti marketing per capitalizzare sul loro momentum.

Raccomandazioni Strategiche:

- **Per entrambi i brand:** Sfruttare il trend di mercato positivo per focalizzare le risorse sulla conversione dell'interesse in vendite.
- **Per Prima Assicurazioni:** Rivedere la strategia pubblicitaria per garantire che gli investimenti generino una crescita incrementale superiore a quella organica.
- **Per Allianz:** Utilizzare la propria posizione di forza per investire proattivamente in innovazione digitale e fidelizzazione, al fine di neutralizzare la minaccia a lungo termine dei disruptor.

2 Contesto di Business e Dati

2.1 Obiettivo dello Studio

Il presente studio si pone un triplice obiettivo:

1. Quantificare l'impatto di una significativa campagna marketing lanciata da Prima Assicurazioni sull'interesse di ricerca generato per il proprio brand (Auto-Impatto).
2. Misurare l'impatto della stessa campagna sull'interesse di ricerca del principale incumbent di mercato, Allianz (Impatto Competitivo).
3. Sviluppare un modello previsionale per delineare le traiettorie di crescita future di entrambi i brand a 12 mesi.

2.2 Il Contesto Competitivo

L'analisi esamina le dinamiche tra due attori emblematici del mercato assicurativo italiano:

- **Allianz SE:** Leader di mercato consolidato, con una forte brand recognition storica.
- **Prima Assicurazioni S.p.A.:** Un'insurtech e challenger digitale, nota per le sue strategie di crescita aggressive e l'intenso uso di marketing a performance.

Questa dinamica "incumbent vs. disruptor" rende l'analisi particolarmente rilevante per comprendere le attuali tendenze del settore.

2.3 L'Evento di Intervento

L'evento catalizzatore dello studio è la campagna pubblicitaria multicanale lanciata da Prima Assicurazioni in data 6 Maggio 2024. La campagna, dal costo dichiarato superiore ai 10 milioni di euro, prevede una durata di 18 mesi e utilizza come testimonial l'attore di fama internazionale Patrick Dempsey.¹

2.4 Fonte Dati e Metrica di Analisi

Per quantificare l'interesse del pubblico in modo oggettivo e tempestivo, è stato utilizzato l'indice di ricerca fornito da **Google Trends**, tale indice misura l'inditeresse su una scala da zero a cento. I dati sono stati raccolti con frequenza mensile per il mercato italiano, coprendo il periodo da Gennaio 2020 a Giugno 2025. Questa metrica serve come un efficace proxy per la "brand awareness" e l'interesse "top-of-funnel", misurando l'intenzione di ricerca degli utenti prima che questa si traduca in una potenziale interazione commerciale.

3 Analisi Esplorativa dei Dati (EDA)

3.1 Visualizzazione Iniziale e Dinamiche di Trend

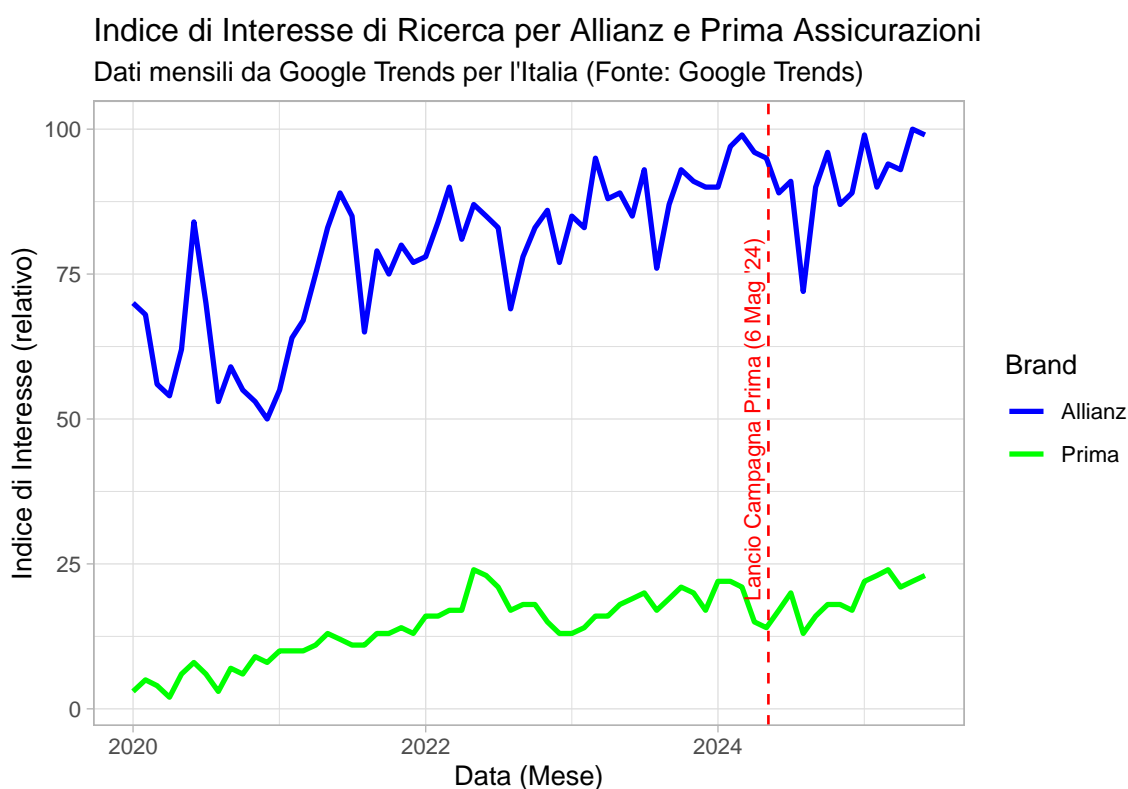


Figure 1: Indice di Interesse Storico Comparativo, Gen 2020 - Giu 2025.

¹Fonte: Engage.it, "PRIMA ASSICURAZIONI PUNTA SU PATRICK DEMPSEY PER LA NUOVA CAMPAGNA PUBBLICITARIA", 6 Maggio 2024. Link: <https://www.engage.it/media/prima-assicurazioni-nuova-campagna-patrick-dempsey>

Table 1: Statistiche Descrittive Principali per Brand.

Brand	Media	St.Dev.	Min	Q1	Mediana	Q3	Max
Allianz	80.76	13.46	50	72	84.5	90	100
Prima Assicurazioni	14.86	5.76	2	11	16.0	19	24

L'analisi visiva preliminare (Figure 1) e le statistiche descrittive (Table 1) confermano la struttura di mercato “incumbent vs. challenger”. Allianz mantiene una dominanza netta nell'interesse di ricerca medio, con un indice che ha raggiunto il valore massimo di 100 nel periodo analizzato.

Tuttavia, l'analisi delle traiettorie di crescita rivela una dinamica più complessa. Mentre Allianz mostra una crescita solida, con un incremento del **43%** sull'intero periodo, Prima Assicurazioni esibisce un momentum eccezionale. Il suo indice di interesse è aumentato del **700%** da Gennaio 2020, un chiaro indicatore quantitativo della sua aggressiva strategia di penetrazione del mercato.

3.2 Decomposizione e Pattern Stagionali

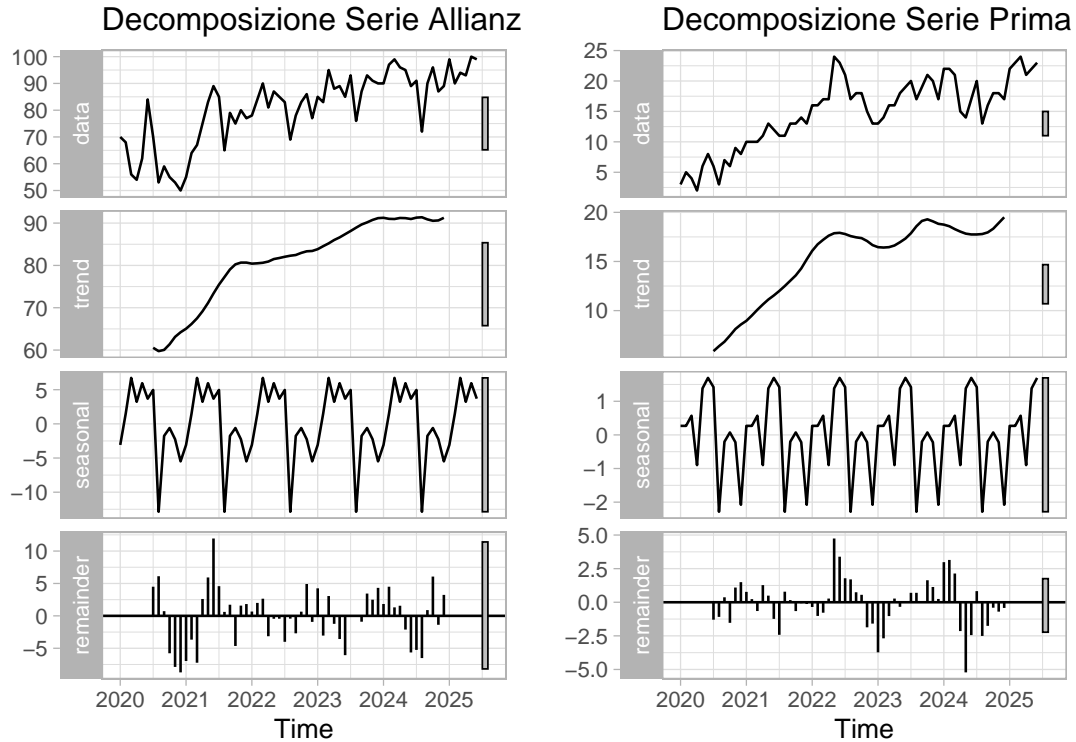


Figure 2: Decomposizione delle Serie Storiche per Brand. Allianz (sinistra), Prima Assicurazioni (destra)

La scomposizione additiva delle serie (Figure 2) permette di isolare e analizzare le componenti di trend e stagionalità.

- **Trend:** Per entrambi i brand, viene confermata una chiara tendenza di fondo positiva, suggerendo un crescente interesse generale verso i prodotti assicurativi in Italia.
- **Stagionalità** L'analisi stagionale rivela pattern comportamentali distinti:
 - **Allianz** esibisce una stagionalità annuale prevedibile e stabile, caratterizzata da un calo sistematico dell'interesse nel mese di agosto, in linea con le dinamiche feriali nazionali.

- **Prima Assicurazioni** mostra una struttura più complessa e distintiva. Emerge un chiaro ciclo trimestrale, non immediatamente visibile nei dati grezzi. Questa scoperta suggerisce l'esistenza di dinamiche di business o di marketing interne che guidano l'interesse a una frequenza più alta rispetto a quella del mercato generale.

La presenza di queste componenti strutturate (trend e stagionalità) rende inadeguata un'analisi basata sui soli dati grezzi e motiva la necessità di ricorrere a modelli di serie storiche specifici, come verrà dettagliato nella sezione successiva.

4 Metodologia

4.1 Approccio Generale: Analisi di Intervento con Modelli ARIMAX

Per quantificare l'impatto della campagna marketing di Prima Assicurazioni è stata adottata una metodologia di **Analisi di Intervento**. Questo approccio statistico è specificamente progettato per isolare e misurare l'effetto di un evento esterno e databile (l'intervento) sulla dinamica di una serie storica.

Il framework tecnico utilizzato è quello della **Regressione con Errori ARIMA**, comunemente noto come modello **ARIMAX**. Questo modello agisce su due livelli:

1. Una componente di **regressione** stima l'impatto dell'intervento, rappresentato da una variabile esogena (la 'X' di ARIMAX). Nel nostro caso, questa è una variabile dummy che assume valore 0 prima della campagna e 1 dopo il suo inizio.
2. Una componente **SARIMA** (Seasonally Adjusted ARIMA) modella la "normale" dinamica interna della serie storica (trend, stagionalità e autocorrelazione).

L'obiettivo è quindi quello di "pulire" la serie da tutte le sue correlazioni interne per far emergere l'effetto puro dell'intervento. Il coefficiente stimato per la variabile di intervento, se statisticamente significativo, fornirà una misura quantitativa dell'impatto della campagna.

4.2 Pre-processing e Test di Stazionarietà

Un prerequisito fondamentale per la modellizzazione ARIMA è la **stazionarietà** della serie storica, ovvero la condizione in cui proprietà statistiche come media e varianza rimangono costanti nel tempo. L'analisi di una serie non stazionaria può portare a stime di modelli inaffidabili e a previsioni errate.

Per verificare formalmente questa condizione, è stato utilizzato il **test di Dickey-Fuller Aumentato (ADF)**. Un p-value del test inferiore alla soglia di significatività ($\alpha = 0.05$) permette di rifiutare l'ipotesi di non-stazionarietà.

Come evidenziato dall'analisi esplorativa, entrambe le serie storiche originali risultano non-stazionarie, risultato confermato dal test ADF (Table 2). È stato quindi necessario un pre-processing dei dati per renderli idonei alla modellizzazione, con strategie differenziate per i due brand:

- **Allianz:** La stazionarietà è stata raggiunta tramite una **differenziazione del primo ordine**, una trasformazione standard per rimuovere il trend lineare.
- **Prima Assicurazioni:** La complessa natura del trend (esponenziale) e della stagionalità ha richiesto un approccio più robusto. Si è optato per una **trasformazione logaritmica** (per stabilizzare la varianza e linearizzare il trend) seguita da una **differenziazione del primo ordine**.

4.3 Identificazione e Selezione dei Modelli

La selezione del modello SARIMAX ottimale per ciascuna serie storica è stata condotta attraverso un processo iterativo e rigoroso, volto a bilanciare l'adeguatezza del fit con la parsimonia del modello. La strategia si è articolata in tre fasi principali:

1. **Identificazione dei parametri Iniziali:** Una prima ipotesi sulla struttura del modello (gli ordini p , q , P , Q) è stata formulata attraverso l'analisi delle funzioni di autocorrelazione (ACF) e autocorrelazione parziale (PACF) delle serie stazionarie.
2. **Stima e validazione diagnostica:** I modelli candidati sono stati quindi stimati e sottoposti a una rigorosa diagnostica dei residui. Un modello è stato considerato "valido" solo se i suoi residui si sono dimostrati indistinguibili da un rumore bianco, ovvero privi di autocorrelazione residua. Questa condizione è stata verificata formalmente tramite il **test di Ljung-Box**, con una soglia di significatività del p -value > 0.05 .
3. **Selezione competitiva del modello finale:** I modelli che hanno superato la fase di validazione sono stati messi a confronto diretto sulla base di due criteri chiave:
 - **Bontà del Fit e Parsimonia (In-Sample):** Misurata tramite i **criteri di informazione di Akaike e Bayesiano (AIC/BIC)**. Valori più bassi indicano un equilibrio migliore tra complessità e fit sui dati di training.
 - **Accuratezza Predittiva (Out-of-Sample):** Misurata tramite il **Root Mean Squared Error (RMSE)** calcolato sulle previsioni a 12 mesi su un set di test non utilizzato durante la stima del modello. Questo criterio è stato considerato decisivo nella scelta finale.

Applicando questo rigoroso processo di stima, validazione e selezione competitiva, è stato possibile identificare il modello ottimale per ciascun brand. La sezione seguente presenta in dettaglio i risultati di questa analisi, a partire dalla selezione dei modelli finali fino alla stima dell'impatto dell'intervento e alle proiezioni previsionali.

5 Risultati e Analisi

5.1 Validazione del Pre-processing e Stazionarietà delle Serie

Table 2: Risultati del Test ADF prima e dopo il Pre-processing.

Serie Storica	Stato dei Dati	p-value Test ADF	Risultato ($\alpha=0.05$)
Allianz	Originale	0.075	Non Stazionaria
Allianz	Trasformata	0.010	Stazionaria
Prima Assicurazioni	Originale	0.228	Non Stazionaria
Prima Assicurazioni	Trasformata	0.014	Stazionaria

Come primo risultato, si valida l'efficacia delle strategie di pre-processing. La Tabella 2 (Table 2) mostra come le serie originali, entrambe non-stazionarie secondo il test ADF, siano state trasformate con successo. L'applicazione dei test sulle serie finali (Grafico 3) conferma il raggiungimento della stazionarietà per entrambi i brand, rendendoli idonei per la successiva fase di modellizzazione.

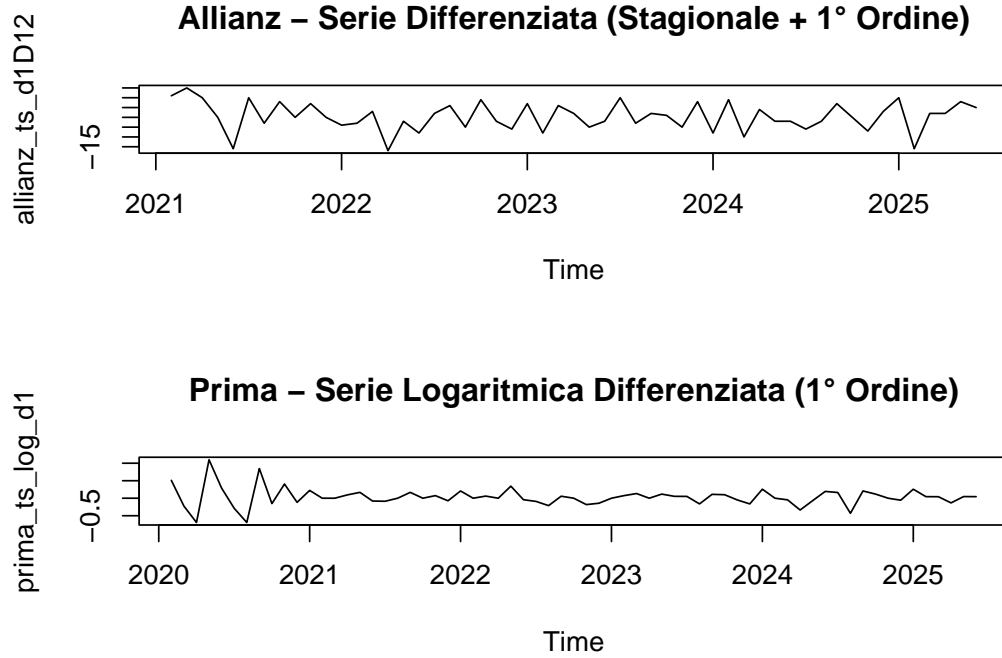


Figure 3: Serie Storiche Trasformate e Stazionarie per Brand.

5.2 Selezione del Modello per Allianz

Table 3: Confronto dei Modelli Candidati per Allianz.

Modello Candidato	AIC	BIC	RMSE (Test Set)	Esito Ljung-Box
A: SARIMA(0,1,0)(0,1,0)[12]	361.8	363.8	6.48	Superato
B: SARIMA(1,1,1)(1,0,0)[12]	439.1	447.8	5.80	Superato

Per la serie Allianz, la selezione ha contrapposto un modello Seasonal Random Walk (Modello A), caratterizzato da massima parsimonia, a un modello SARIMA(1,1,1)(1,0,0)[12] (Modello B), più complesso ma flessibile.

Come illustrato in Tabella 3, sebbene il Modello A presenti criteri di informazione (AIC/BIC) nettamente inferiori, il **criterio decisivo dell'accuratezza predittiva ha favorito il Modello B**, che ha registrato un Root Mean Squared Error (RMSE) sul test set inferiore. Questa superiore capacità previsionale, unita a residui di qualità leggermente migliore, ha determinato la sua selezione come modello finale. L'analisi diagnostica del modello scelto (Figure 4) conferma l'assenza di autocorrelazione residua.

Table 4: Coefficienti Stimati per il Modello Finale di Allianz (SARIMA(1,1,1)(1,0,0)[12]).

Termine del Modello	Stima del Coefficiente	Errore Standard
AR(1) - ar1	0.6658	0.1609
MA(1) - ma1	-0.9268	0.0693
SAR(1) - sar1	0.6369	0.1107

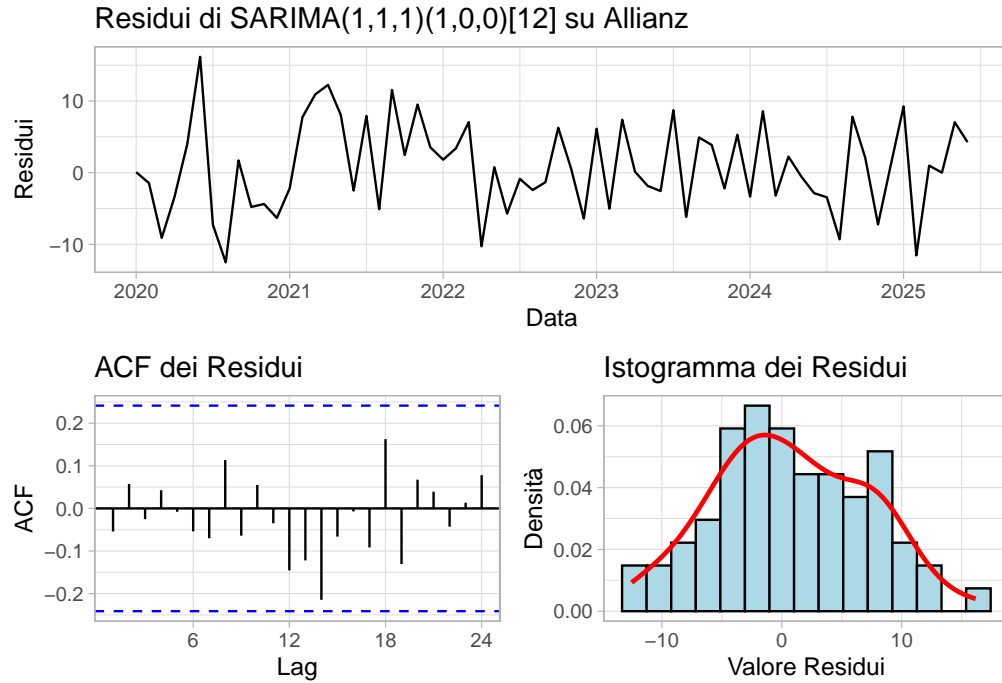


Figure 4: Analisi dei Residui del modello finale per Allianz

5.3 Selezione del Modello per Prima Assicurazioni

Table 5: Confronto dei Modelli Candidati per Prima Assicurazioni (su serie logaritmica).

Modello Candidato	AIC	BIC	RMSE (Test Set)	Esito Ljung-Box
A: ARIMA(4,1,0) with drift	-17.28	-4.23	2.33	Superato ($p=0.072$)
B: ARIMA(3,1,0)	-0.69	8.01	4.55	Superato ($p=0.543$)

La selezione del modello per Prima Assicurazioni ha richiesto un'analisi più approfondita, data la complessa natura del trend della serie. La strategia finale si è basata su una trasformazione logaritmica dei dati, con il confronto tra due modelli ARIMA candidati (Table 5).

Nonostante il Modello B presenti residui tecnicamente superiori, il Modello A è risultato **nettamente vincente su due fronti critici**:

1. Una **bontà di fit in-sample** drasticamente migliore, come indicato dai valori di AIC e BIC inferiori.
2. Un'**accuratezza predittiva out-of-sample** quasi doppia, con un RMSE di **2.33** contro i **4.54** del modello concorrente.

Il potere predittivo superiore ha reso il Modello B la scelta inequivocabile per l'analisi. La diagnostica del modello (Figure 5) mostra residui che, sebbene non perfetti, superano il test di Ljung-Box, confermandone la validità.

Table 6: Coefficienti Stimati per il Modello Finale di Prima (ARIMA(4,1,0) with drift su log).

Termine del Modello	Stima del Coefficiente	Errore Standard
AR(1) - ar1	-0.3286	0.1149
AR(2) - ar2	-0.3466	0.1126
AR(3) - ar3	-0.2192	0.1187
AR(4) - ar4	0.5070	0.1246
Drift	0.0326	0.0169

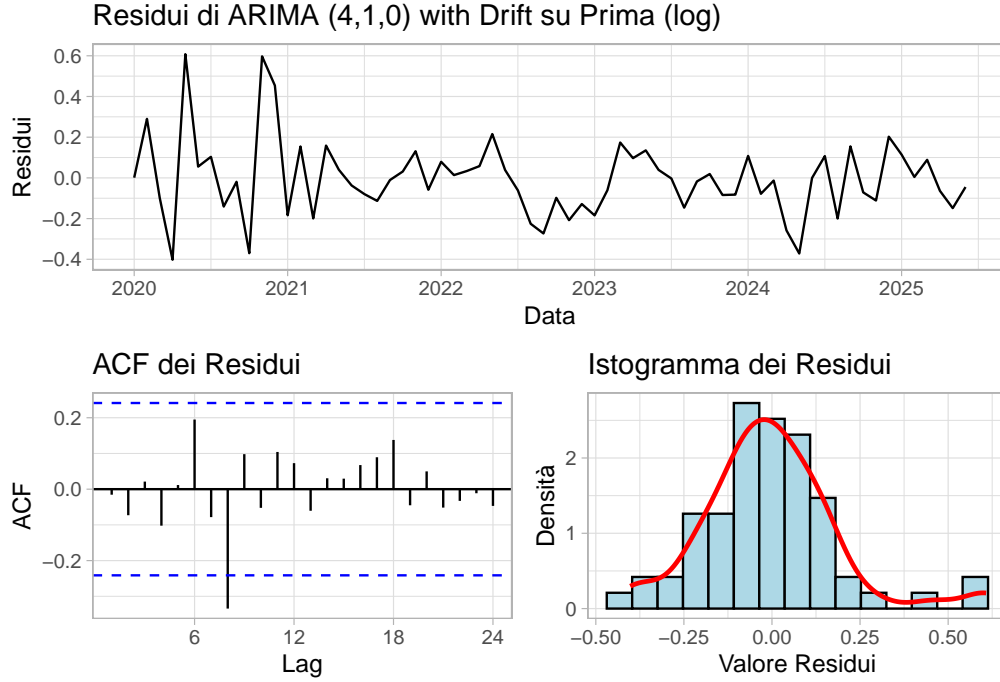


Figure 5: Analisi dei Residui del modello finale per Prima

5.4 Stima dell'Impatto della Campagna

Utilizzando i modelli ARIMAX finali, è stato possibile quantificare l'impatto della campagna marketing sia sul brand che l'ha lanciata, sia sul concorrente.

5.4.1 Auto-Impatto: Un Rallentamento della Crescita per Prima Assicurazioni

Table 7: Risultati del Modello di Intervento per Prima Assicurazioni.

Coefficiente	Stima	Errore.Std.	t.statistic	p.value..approx..
xreg (log)	-0.3137	0.149	-2.105	< 0.05

L'analisi dell'impatto sulla serie di Prima Assicurazioni rivela un risultato contro-intuitivo e di alto valore

strategico. Come mostra la Tabella 7, il coefficiente associato all'intervento risulta **negativo e statisticamente significativo**.

Sebbene i dati grezzi mostrino una crescita dell'interesse di ricerca dopo il lancio della campagna, il modello stima che, una volta scontato il fortissimo trend pre-esistente, l'impatto netto dell'intervento sia stato un **rallentamento della crescita pari al 27%** rispetto alla traiettoria attesa. Questo suggerisce che la campagna, pur generando visibilità, si sia dimostrata meno efficiente nel guidare l'interesse incrementale rispetto al momentum "organico" che il brand stava già sperimentando.

5.4.2 Impatto Competitivo: Resilienza del Brand Allianz

Table 8: Risultati del Modello di Intervento per Allianz.

Coefficiente	Stima	Errore.Std.	t.statistic	p.value..approx..
xreg	-4.777	5.0212	-0.951	> 0.05

L'analisi dell'impatto sul leader di mercato Allianz porta a una conclusione differente. Il modello rileva un leggero calo nell'interesse di ricerca, ma il coefficiente associato all'intervento **non risulta statisticamente significativo** (Table 8).

Questo risultato indica che la massiccia campagna pubblicitaria del competitor non è riuscita a erodere in modo misurabile la "share of mind" online del brand Allianz. Ciò suggerisce una notevole **resilienza del brand** e una solida posizione di mercato, che non viene facilmente intaccata da offensive di marketing anche su larga scala.

5.5 Analisi Previsionale Comparativa

Il grafico finale (Figure 6) sintetizza l'analisi storica e proietta le traiettorie future dei due brand per i prossimi 12 mesi. L'elemento più significativo che emerge dalle previsioni è la **potenziale e rapida riduzione del divario** tra il leader di mercato e il suo challenger digitale.

- **Traiettoria di Prima Assicurazioni:** Il modello per Prima proietta una **forte crescita accelerata**, proseguendo il momentum esponenziale osservato nei dati storici. Se questa tendenza venisse confermata, l'interesse di ricerca per il brand potrebbe avvicinarsi notevolmente a quello del leader di mercato entro il prossimo anno. Gli intervalli di confidenza, sebbene ampi, confermano un'elevata probabilità di crescita sostenuta.
- **Traiettoria di Allianz:** Per Allianz, il modello prevede una **crescita più matura e stabile**. La previsione suggerisce il mantenimento di un elevato livello di interesse di ricerca, ma con un tasso di crescita modesto. Questa stabilità, pur confermando la solida posizione del brand, lo espone al rischio di essere raggiunto da competitor più agili e con tassi di crescita superiori.

In sintesi, l'analisi previsionale quantifica uno scenario di mercato dinamico, in cui la posizione dominante di Allianz, se non supportata da strategie proattive, potrebbe essere messa in discussione nel medio termine dal rapido slancio dei nuovi attori digitali.

6 Interpretazione di Business e Raccomandazioni Strategiche

L'analisi statistica ha prodotto insight di chiara rilevanza strategica, traducibili in raccomandazioni operative per i brand analizzati. Le conclusioni si articolano attorno a tre temi chiave: la dinamica generale del mercato, l'efficacia dell'investimento marketing di Prima e le traiettorie competitive future.

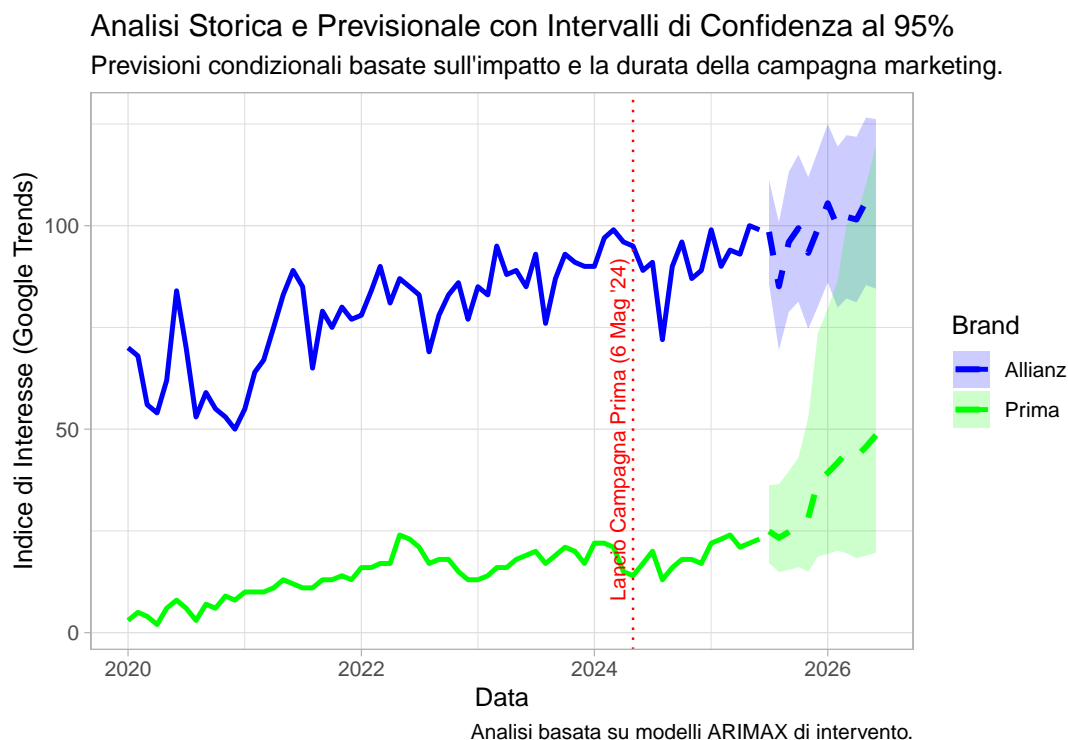


Figure 6: Previsioni a 12 Mesi per Brand con intervalli di Confidenza al 95%.

6.1 Il Vento a Favore: Sfruttare il Momentum del Mercato

L'analisi del trend ha evidenziato un contesto di mercato favorevole, con un interesse di ricerca online per i prodotti assicurativi in crescita sostenuta per entrambi i brand. Le previsioni indicano che questo “tailwind” di mercato proseguirà nel medio termine.

- **Raccomandazione Strategica:** Per entrambi i brand, l'imperativo è capitalizzare su questo crescente interesse. Si raccomanda di allocare risorse per ottimizzare il tasso di conversione lungo il funnel di marketing (dall'interesse all'azione), al fine di massimizzare l'acquisizione di clienti in questa fase di espansione del mercato.

6.2 Efficacia ed Efficienza della Campagna: Un Insight Contro-Intuitivo

L'analisi di intervento ha rivelato che la massiccia campagna pubblicitaria di Prima Assicurazioni è stata, nel periodo analizzato, sia **inefficace** nel suo impatto competitivo sia **inefficiente** rispetto al proprio momentum organico.

- **Implicazioni per Prima Assicurazioni:** Il risultato suggerisce la necessità di una revisione strategica delle future campagne. L'investimento non ha generato un'accelerazione della crescita superiore a quella già in atto, indicando un potenziale margine per ottimizzare il budget di marketing verso iniziative a più alto rendimento incrementale.
- **Implicazioni per Allianz:** L'assenza di un impatto negativo significativo sul proprio brand è un segnale di forza. Ciò suggerisce che non sia necessaria una reazione diretta tramite una costosa contro-campagna. Le risorse possono essere allocate su iniziative a più alto valore aggiunto, come quelle delineate nel punto seguente.

6.3 Lo Scenario Futuro: Il Divario si Restringe

Le proiezioni a 12 mesi indicano un mercato altamente dinamico. Il forte momentum di Prima Assicurazioni, se mantenuto, potrebbe portare a una significativa riduzione del divario con il leader di mercato.

- **Prospettiva per Prima Assicurazioni:** La traiettoria di crescita è eccezionale. La strategia futura dovrebbe focalizzarsi sul mantenimento di questo slancio attraverso un marketing aggressivo (ma, come visto, più efficiente) e una continua innovazione di prodotto.
- **Prospettiva per Allianz:** Sarebbe strategicamente imprudente ignorare la velocità di crescita dei challenger digitali. La raccomandazione è di adottare una strategia proattiva di difesa della leadership, non tanto nel campo della pubblicità tradizionale, quanto nell'**innovazione digitale** di prodotti e servizi e in iniziative mirate alla **fidelizzazione (retention)** della base clienti esistente.

7 Limiti e Sviluppi Futuri

Come ogni analisi basata su dati pubblici, questo studio presenta alcuni limiti intrinseci che aprono la strada a futuri approfondimenti.

7.1 Limiti Metodologici

- **Dati Proxy:** L'analisi si basa sull'indice di Google Trends, un eccellente proxy per l'interesse di ricerca "top-of-funnel". Tuttavia, non misura direttamente le conversioni. Un'analisi completa del ROI finanziario richiederebbe l'integrazione di dati proprietari di vendita.
- **Validità del Modello Prima:** Il modello finale per Prima Assicurazioni, sebbene predittivamente superiore, ha mostrato residui che superano i test di validità statistica solo marginalmente. Si raccomanda un monitoraggio continuo e un potenziale refining del modello con l'acquisizione di nuovi dati futuri.

7.2 Prospettive di Sviluppo Futuro

I risultati di questo studio suggeriscono un'importante area di ricerca strategica per qualsiasi attore del mercato assicurativo:

- **Analisi di Benchmark Settoriale (Insurtech vs. Incumbent):** Lo studio potrebbe essere esteso a un paniere più ampio di brand, includendo diverse insurtech e compagnie tradizionali. Questo permetterebbe di:
 - Creare un **benchmark quantitativo** per misurare la velocità con cui i nuovi player digitali stanno guadagnando "share of mind" rispetto agli incumbent.
 - Identificare i **driver di crescita** comuni tra le insurtech di successo.
 - Sviluppare **indicatori di allerta precoce (early warning indicators)** per intercettare nuove minacce competitive e interpretare le future dinamiche di mercato.

Appendice A: Dettagli Metodologici e Diagnostica

A.1 Grafici di Autocorrelazione (ACF/PACF) delle serie stazionarie

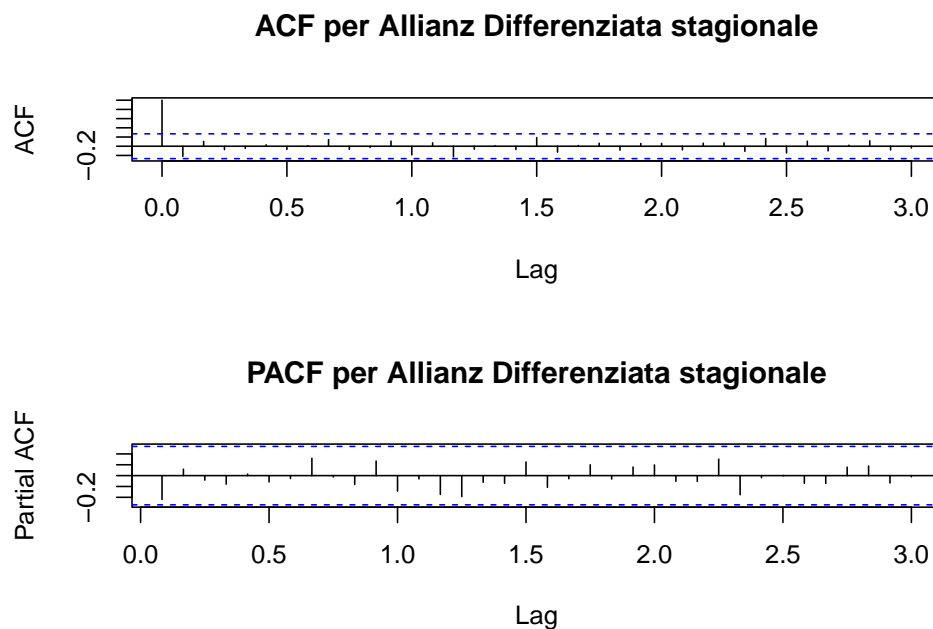


Figure 7: A.1 Autocorrelogrammi Allianz stagionaria, indicazione di un possibile Seasonal Random Walk

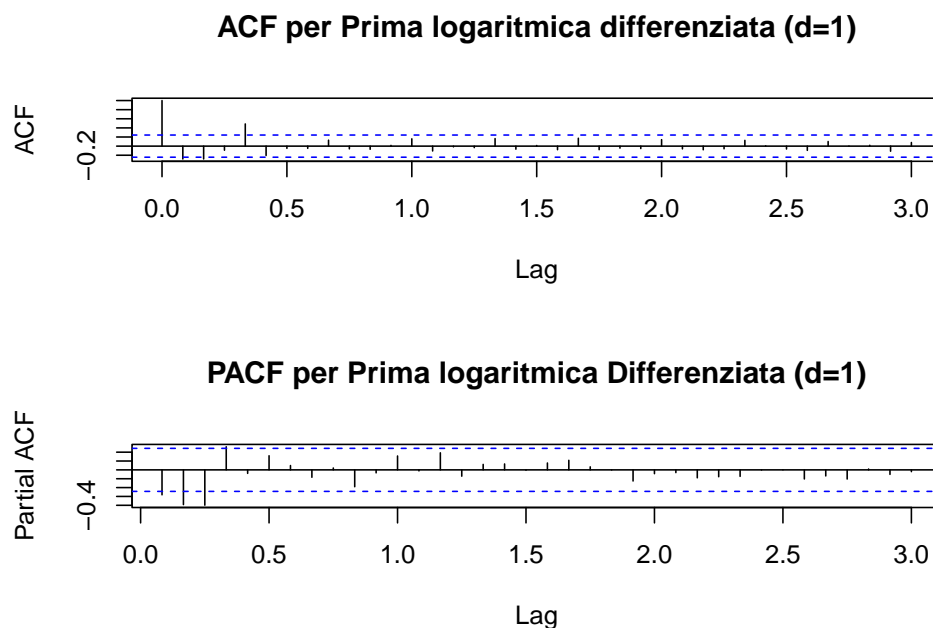


Figure 8: A.1 Autocorrelogrammi Prima logaritmica stagionaria, si teorizzava un ARIMA(3,1,0)

A.2 Output di Stima e Diagnostica dei Modelli Candidati per Allianz

A.2.1 Modello A Scartato Allianz: SARIMA(0,1,0)(0,1,0)[12]

Questo modello (Seasonal Random Walk) non ha coefficienti AR o MA da stimare. La sua performance è valutata unicamente sulle statistiche di sintesi.

Statistiche di Sintesi del Modello

Table 9: A.2.1 Statistiche di bontà del fit per il modello scartato di Allianz.

sigma	logLik	AIC	BIC	nobs
7.21	-179.9	361.8	363.77	53

Analisi dei Residui del Modello

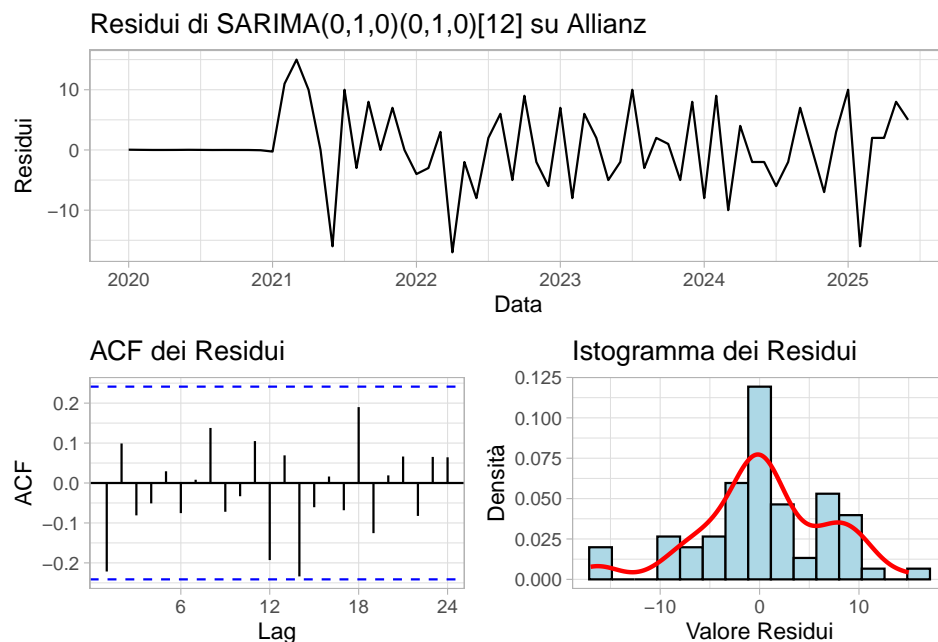


Figure 9: A.2.1 Analisi dei Residui del modello scartato per Allianz

QQ-plot del Modello

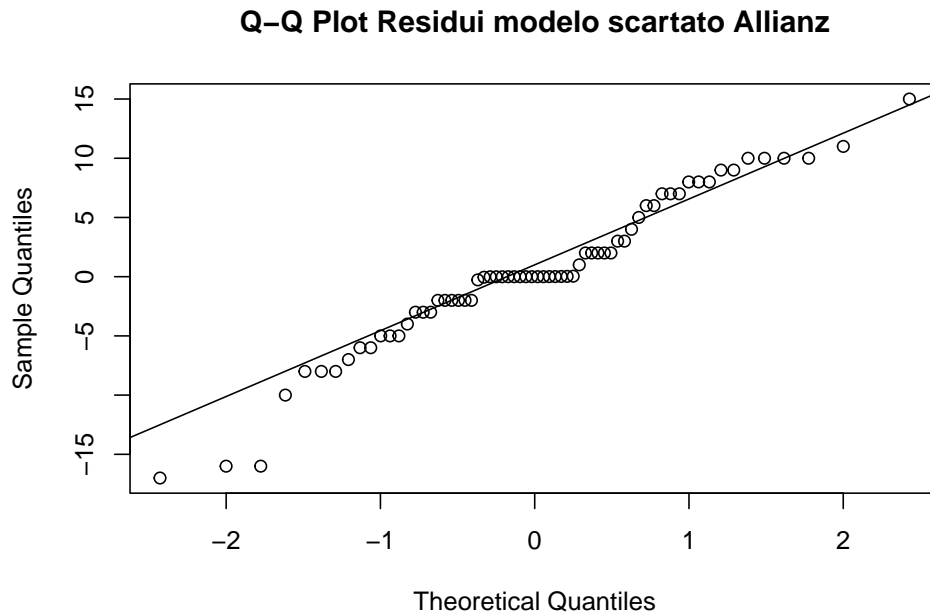


Figure 10: A.2.1 QQ-Plot dei residui del modello scartato per Allianz.

A.2.2 Modello B Vincitore Allianz: SARIMA(1,1,1)(1,0,0)[12]

Tabella dei Coefficienti

Table 10: A.2.2 Coefficienti stimati, errori standard e statistiche per il modello finale di Allianz.

term	estimate	std.error
ar1	0.6658	0.1609
ma1	-0.9268	0.0693
sar1	0.6369	0.1107

Statistiche di Sintesi del Modello

Table 11: A.2.2 Statistiche di bontà del fit per il modello finale di Allianz.

sigma	logLik	AIC	BIC	nobs
6.51	-215.57	439.14	447.84	65

QQ-Plot del Modello

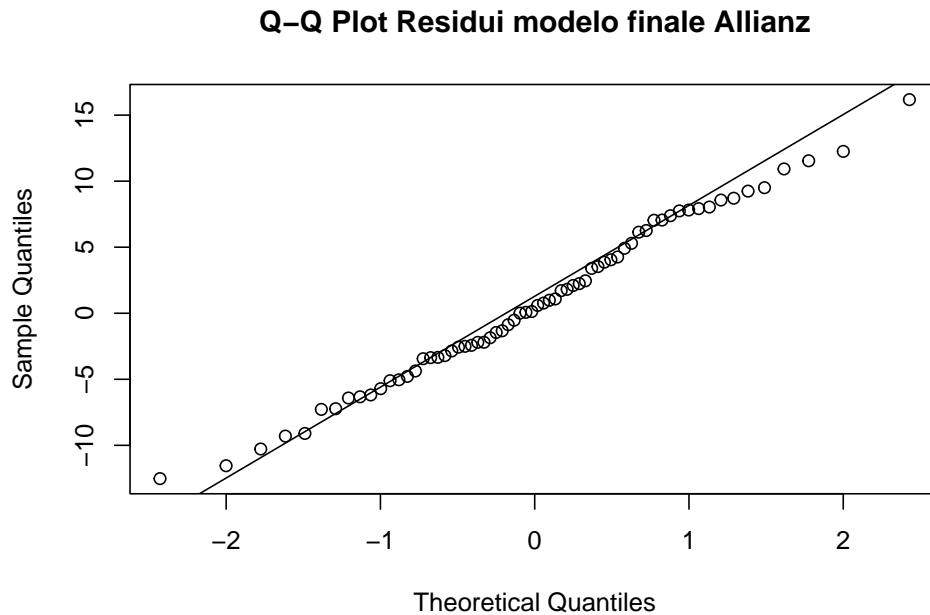


Figure 11: A.2.2 QQ-Plot dei residui del modello finale per Allianz.

A.3 Output di Stima e Diagnostica dei Modelli Candidati per Prima Assicurazioni

A.3.1 Modello A Vincitore Prima: ARIMA(4,1,0) with Drift

Tabella dei Coefficienti

Table 12: A.3.1 Coefficienti stimati, errori standard e statistiche per il modello finale di Prima.

term	estimate	std.error
ar1	-0.3286	0.1149
ar2	-0.3466	0.1126
ar3	-0.2192	0.1187
ar4	0.5070	0.1246
drift	0.0326	0.0169

Statistiche di Sintesi del Modello

Table 13: A.3.1 Statistiche di bontà del fit per il modello finale di Prima.

sigma	logLik	AIC	BIC	nobs
0.2	14.64	-17.28	-4.23	65

QQ-Plot del Modello

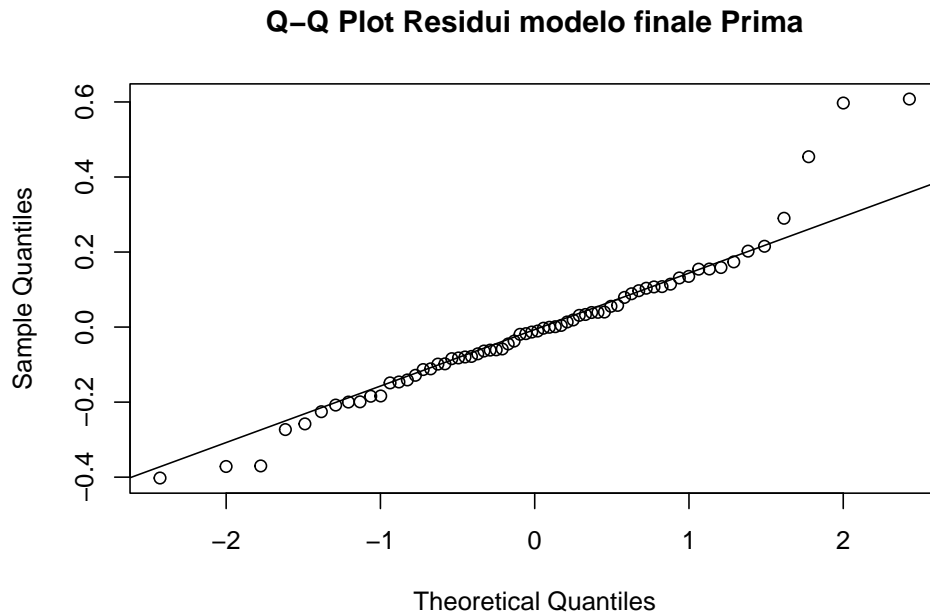


Figure 12: A.3.1 QQ-Plot dei residui del modello finale per Prima.

A.3.2 Modello B scartato Prima: ARIMA(3,1,0)

Tabella dei Coefficienti

Table 14: A.3.2 Coefficienti stimati, errori standard e statistiche per il modello scartato di Prima.

term	estimate	std.error
ar1	-0.4872	0.1166
ar2	-0.5073	0.1180
ar3	-0.3546	0.1227

Statistiche di Sintesi del Modello

Table 15: A.3.2 Statistiche di bontà del fit per il modello scartato di Prima.

sigma	logLik	AIC	BIC	nobs
0.23	4.34	-0.69	8.01	65

Analisi dei Residui del Modello

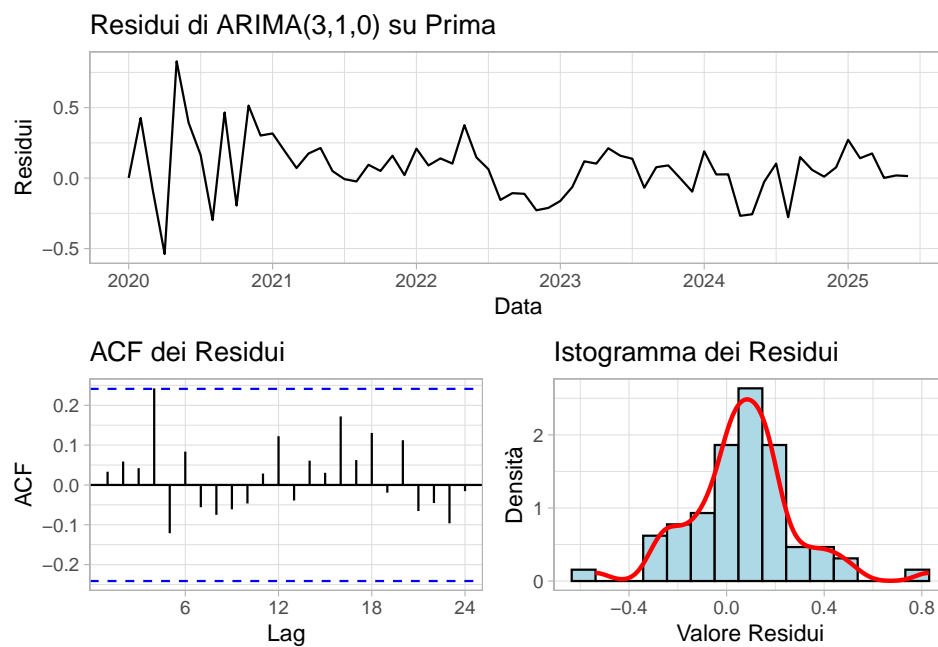


Figure 13: A.3.2 Analisi dei Residui del modello scartato per Prima

QQ-Plot del Modello

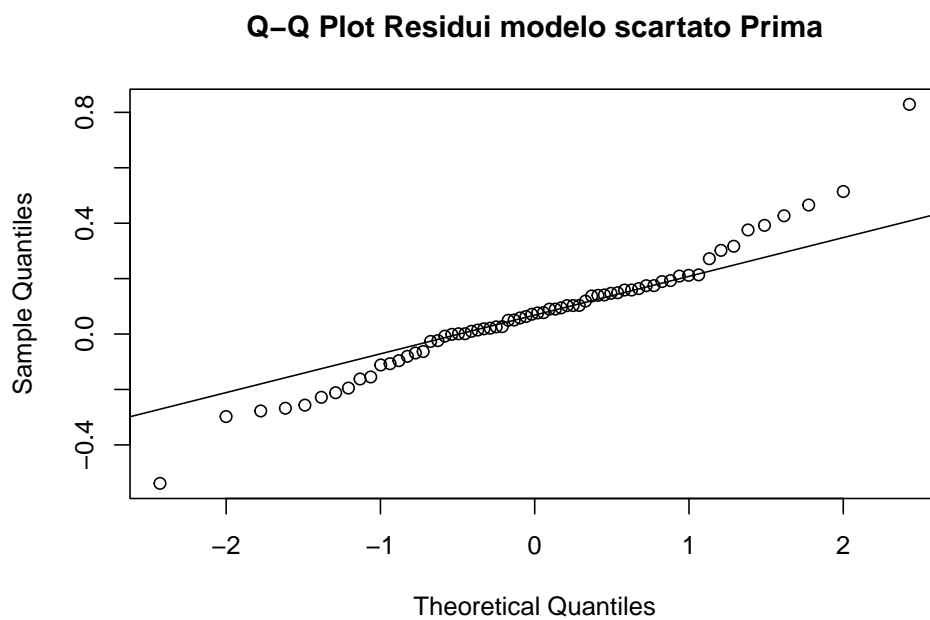


Figure 14: A.3.2 QQ-Plot dei residui del modello scartato per Prima.

A.4 Output dei Modelli di Intervento Finali

A.4.1 Modello ARIMAX Allianz

Tabella dei Coefficienti

Table 16: A.4.1 Coefficienti stimati, errori standard e statistiche per il modello ARIMAX di Allianz.

term	estimate	std.error
ar1	0.6299	0.1785
ma1	-0.9087	0.0861
sar1	0.6438	0.1100
xreg	-4.7770	5.0212

Statistiche di Sintesi del Modello

Table 17: A.4.1 Statistiche di bontà del fit per il modello ARIMAX di Allianz.

sigma	logLik	AIC	BIC	nobs
6.5	-215.13	440.26	451.13	65

A.4.2 Modello ARIMAX Prima

Tabella dei Coefficienti

Table 18: A.4.2 Coefficienti stimati, errori standard e statistiche per il modello ARIMAX di Prima.

term	estimate	std.error
ar1	-0.3473	0.1166
ar2	-0.3716	0.1137
ar3	-0.2679	0.1206
ar4	0.4956	0.1258
drift	0.0374	0.0154
xreg	-0.3137	0.1490

Statistiche di Sintesi del Modello

Table 19: A.4.2 Statistiche di bontà del fit per il modello ARIMAX di Prima.

sigma	logLik	AIC	BIC	nobs
0.19	16.74	-19.48	-4.26	65

Appendice B: Codice R Completo e Commentato

```
# =====
# Progetto: Analisi di Intervento Allianz vs. Prima
# Autore: Valerio Lapiello
# Data: 19/07/2025
# =====

# -----
# 1. SETUP E CARICAMENTO DEI DATI
# -----

# Caricamento delle librerie necessarie

library(readr) # Per la lettura efficiente dei file CSV (read_csv)
library(here) # Per la gestione robusta dei percorsi dei file
library(tidyverse) # Per manipolare e visualizzare i dati
library(summarytools) # Per le statistiche descrittive
library(ggfortify) # Per estendere ggplot2 a oggetti di altri pacchetti (es. decompose)
library(patchwork) # Per combinare facilmente più grafici ggplot2 in un'unica figura
library(broom) # Per convertire i summary in tabelle
library(tseries) # Per i test di stazionarietà (adf.test)
library(forecast) # Per la modellizzazione e previsione (Arima, auto.arima)
library(zoo) # Per la gestione di oggetti di tipo data/ora

# Caricamento e visualizzazione Dati grezzi
dati_raw <- read_csv(here("data", "multiTimeline.csv"),
                     col_types = cols(Mese = col_date(format = "%Y-%m")),
                     skip = 2)

View(dati_raw)

# Rinomina colonne e pulizia
colnames(dati_raw)[2] <- "Allianz"
colnames(dati_raw)[3] <- "Prima"

dati_clean <- dati_raw %>%
  mutate_at(vars(Allianz, Prima), function(x) {
    as.numeric(str_replace(x, "<1", "0"))
  })
# Controllo primi dati
head(dati_clean)

# -----
# 2. RAPPRESENTAZIONE GRAFICA
# -----

# Trasformazione dati per ggplot2
dati_plot <- dati_clean %>%
  pivot_longer(
    cols = c("Allianz", "Prima"),
    names_to = "Brand",
    values_to = "IndiceRicerca"
  )
```

```

# Grafico comparativo
ggplot(dati_plot, aes(x = Mese, y = IndiceRicerca, color = Brand)) +
  geom_line(linewidth = 1) +
  geom_vline(xintercept = as.Date("2024-05-06"), linetype = "dashed", color = "red") +
  annotate("text", x = as.Date("2024-05-06"), y = 50,
    label = "Lancio Campagna Prima (6 Mag '24)",
    color="red", angle=90, vjust = -0.5, size=3) +
  labs(
    title = "Indice di Interesse di Ricerca per Allianz e Prima Assicurazioni",
    subtitle = "Dati mensili da Google Trends per l'Italia (Fonte: Google Trends)",
    x = "Data (Mese)",
    y = "Indice di Interesse (relativo)",
    color = "Brand"
  ) +
  scale_color_manual(values = c("Allianz" = "blue", "Prima" = "green")) +
  theme_light()

# -----
# 3. DECOMPOSIZIONE E TEST DI STAZIONARIETA' INIZIALE
# -----

# Creazione serie storiche ts
start_year <- as.numeric(format(min(dati_clean$Mese), "%Y"))
start_month <- as.numeric(format(min(dati_clean$Mese), "%m"))

allianz_ts <- ts(
  dati_clean$Allianz,
  start = c(start_year, start_month),
  frequency = 12
)

prima_ts <- ts(
  dati_clean$Prima,
  start = c(start_year, start_month),
  frequency = 12
)

print(allianz_ts) # Controllo

# Statistiche descrittive
allianz_stats = descr(allianz_ts, stats = c("mean", "fivenum", "sd"))
prima_stats = descr(prima_ts, stats = c("mean", "fivenum", "sd"))
print(allianz_stats)
print(prima_stats)

# Decomposizione serie storiche
decomposizione_allianz <- decompose(allianz_ts, type = "additive")
decomposizione_prima <- decompose(prima_ts, type = "additive")

# Grafico Comparativo serie storiche decomposte
plot_decomp_allianz <- autoplot(decomposizione_allianz,
  main = "Decomposizione Serie Allianz") +
  theme_light() # Plot serie decomposta Allianz

```

```

plot_decomp_prima <- autoplot(decomposizione_prima,
                             main = "Decomposizione Serie Prima") +
  theme_light() # Plot serie decomposta Prima
# Rappresentazione affiancata dei due plot
plot_decomp_allianz | plot_decomp_prima

# -----
# 4. PRE-PROCESSING E TEST DI STAZIONARIETA'
# -----

# Differenziazione di primo ordine per entrambe le serie storiche
allianz_ts_d1 <- diff(allianz_ts, differences = 1)
prima_ts_d1 <- diff(prima_ts, differences = 1)

# Grafici per entrambe le serie storiche differenziate 1° Ordine
par(mfrow=c(2,1))
plot(allianz_ts_d1, main = "Allianz - Serie Differenziata (1° Ordine)")
plot(prima_ts_d1, main = "Prima - Serie Differenziata (1° Ordine)")
par(mfrow=c(1,1))

# Test ADF per entrambe le serie storiche (superato se p-value < 0.05)
adf.test(allianz_ts_d1) # p-value < 0.05
adf.test(prima_ts_d1) # p-value ancora > 0.05 passiamo a differenziazione stagionale

# Allianz risulta stazionaria con differenziazione di primo ordine ma proviamo anche
# la differenziazione stagionale data la presenza di forte stagionalità

# Differenziazione Stagionale per entrambe le serie storiche
allianz_ts_d1D12 <- diff(diff(allianz_ts, lag = 12), differences = 1)
prima_ts_d1D12 <- diff(diff(prima_ts, lag = 12), differences = 1)

# Grafici per entrambe le serie storiche differenziate stagionali
plot(allianz_ts_d1D12, main= "Allianz - Serie Differenziata (Stagionale + 1° Ordine)")
plot(prima_ts_d1D12, main = "Prima - Serie Differenziata (Stagionale + 1° Ordine)")

# Test ADF per entrambe le serie storiche (superato se p-value < 0.05)
adf.test(allianz_ts_d1D12)
# p-value < 0.05 usiamo questa data la stagionalità marcata rilevata
adf.test(prima_ts_d1D12)
# il p-value è aumentato, forse la stagionalità influisce meno del previsto
# si passa a differenziazione di secondo ordine

# Differenziazione di secondo ordine per Prima Assicurazioni
prima_ts_d2 <- diff(prima_ts, differences = 2)

# Grafico per Prima differenziata di secondo ordine
plot(prima_ts_d2, main = "Prima - Serie Differenziata (2° Ordine)")

# Test ADF per Prima differenziata secondo ordine (superato se p-value < 0.05)
adf.test(prima_ts_d2) # p-value < 0.05. Il 2° ordine è compatibile con il trend marcato

# Risultato finale Pre-processing:

```



```

# --> Differenziazione stagionale di primo ordine per Allianz, serie allianz_ts_d1D12
# --> Differenziazione di secondo ordine per Prima Assicurazioni, serie prima_ts_d2

# -----
# 5. IDENTIFICAZIONE DEL MODELLO ARIMA/SARIMA PER ALLIANZ
# -----

# Analisi Autocorrelogrammi per Allianz (stazionaria d=1, D=1)
par(mfrow=c(2,1))
acf(allianz_ts_d1D12, main="ACF per Allianz Differenziata", lag.max = 36)
pacf(allianz_ts_d1D12, main="PACF per Allianz Differenziata", lag.max = 36)
par(mfrow=c(1,1))
# Nessun lag significativo il modello sembra essere un Seasonal Random Walk
# SARIMA(0,1,0)(0,1,0)[12]

# Modellizzazione "manuale" Allianz SARIMA(0,1,0)(0,1,0)[12]
modello_manuale_allianz <- Arima(allianz_ts, order=c(0,1,0), seasonal=c(0,1,0))
print(summary(modello_manuale_allianz)) # Summary del modello
# Analisi Residui (Test Ljung-Box superato se p-value > 0.05)
checkresiduals(modello_manuale_allianz) # Diagnostica Superata

# Modellizzazione Allianz con auto.arima
modello_auto_allianz <- auto.arima(
  allianz_ts,
  seasonal = TRUE,
  stepwise = FALSE,
  approximation = FALSE
)
print(summary(modello_auto_allianz)) # Summary del modello
# Analisi Residui (Test Ljung-Box superato se p-value > 0.05)
checkresiduals(modello_auto_allianz) # Diagnostica Superata
# Il modello stimato è un SARIMA(1,1,1)(1,0,0)[12]

# 5.1 CONFRONTO MODELLI ALLIANZ -----

# Si creano due modelli A e B per il confronto
modello_allianz_A <- modello_manuale_allianz # SARIMA(0,1,0)(0,1,0)[12]
modello_allianz_B <- modello_auto_allianz # SARIMA(1,1,1)(1,0,0)[12]

# Confronto modelli (Coeff e AIC e BIC, si preferiscono i valori più bassi)
print(summary(modello_allianz_A))
print(summary(modello_allianz_B))
# AIC BIC MOLTO minori per modello A (gap > 10)

# Confronto distribuzione dei residui rispetto a una Normale mediante QQ-Plot
par(mfrow=c(1,2))
# QQ plot Modello A
qqnorm(residuals(modello_allianz_A), main="Q-Q Plot Residui Allianz: Modello A")
qqline(residuals(modello_allianz_A))
# QQ plot Modello B
qqnorm(residuals(modello_allianz_B), main="Q-Q Plot Residui Allianz: Modello B")
qqline(residuals(modello_allianz_B))
par(mfrow=c(1,1))

```

```

# QQ-plot mostra leggermente migliore il Modello B

# Confronto dati Out-of-Sample
# Teniamo gli ultimi 12 mesi come set di test
train_set_allianz <- window(allianz_ts,
                             end = c(start_year + length(allianz_ts)/12 - 2, 12))
test_set_allianz <- window(allianz_ts,
                           start = c(start_year + length(allianz_ts)/12 - 1, 1))

# Ristimiamo i modelli solo sul training set
modello_allianz_A_train <- Arima(train_set_allianz, model=modello_allianz_A)
modello_allianz_B_train <- Arima(train_set_allianz, model=modello_allianz_B)

# Generiamo le previsioni per i prossimi 12 mesi
forecast_allianz_A <- forecast(modello_allianz_A_train, h=12)
forecast_allianz_B <- forecast(modello_allianz_B_train, h=12)

# Confrontiamo l'accuratezza delle previsioni rispetto ai dati reali del test set
# La metrica più importante qui è RMSE (Root Mean Squared Error),
# si preferiscono i valori più bassi
print("Accuratezza Modello A (Manuale)")
accuracy(forecast_allianz_A, test_set_allianz)

print("Accuratezza Modello B (auto.arima)")
accuracy(forecast_allianz_B, test_set_allianz)
# La Previsione è migliore nel Modello B

# CONCLUSIONE CONFRONTO -
# Per Allianz si sceglie il modello B : SARIMA(1,1,1)(1,0,0)[12]

# -----
# 7. Identificazione del Modello ARIMA/SARIMA per Prima
# -----

# Analisi Autocorrelogrammi per Prima (stazionaria d=2)
par(mfrow=c(2,1))
acf(prima_ts_d2, main="ACF per Prima Differenziata (d=2)", lag.max=36)
pacf(prima_ts_d2, main="PACF per Prima Differenziata (d=2)", lag.max=36)
par(mfrow=c(1,1))
# PACF lag 3 poi "off"; ACF lag 1 e stagionale ogni 4 lag
# si ipotizza componente AR(3) ed MA(1), si nota stagionalità trimestrale
# SARIMA(3,2,1)(0,0,1)[4]

# Modellizzazione Prima Assicurazioni con auto.arima
modello_auto_prima <- auto.arima(
  prima_ts,
  seasonal = TRUE,
  stepwise = FALSE,
  approximation = FALSE
)
print(summary(modello_auto_prima)) # Output del modello
# il summary del modello indica una d=1 ma con d=1 la serie non è stazionaria
# Rifiuto questo modello

```

```

# Modellizzazione "manuale" SARIMA(3,2,1)(0,0,1)[4]
modello_manuale_prima <- Arima(
  prima_ts,
  order = c(3, 2, 1),
  seasonal = list(order = c(0, 0, 1), period = 4)
) # il modello risulta esplosivo si necessita di ridurre i coefficienti

# Modellizzazione "manuale" SARIMA(2,2,1)(0,0,1)[4]
# Utilizzo lo stesso nome dato che il modello precedente non è stato creato
modello_manuale_prima <- Arima(
  prima_ts,
  order = c(2, 2, 1),
  seasonal = list(order = c(0, 0, 1), period = 4)
)
print(summary(modello_manuale_prima)) # Output del modello
# Analisi Residui (Test Ljung-Box superato se p-value > 0.05)
checkresiduals(modello_manuale_prima) # Diagnostica non superata

# La modellizzazione risulta ostica, si prova un approccio "semi-automatico"

# Modellizzazione con auto.arima vincolata a d=2 e stagionalità trimestrale m=4
modello_auto_prima <- auto.arima(
  prima_ts,
  d = 2,
  seasonal.period = 4,
  seasonal = TRUE,
  stepwise = FALSE,
  approximation = FALSE,
  allowdrift = FALSE # Aggiungiamo questo per modelli con d=2
)
# Il modello è troppo complesso per essere stimato,
# bisogna agire su d=2 (trend) o su m=4 (stagionalità)
# Si sceglie di tornare alla fase di pre-processing per la stazionarizzazione
# Si decide di agire sul trend esponenziale utilizzando i logaritmi

# Differenziazione logaritmica 1° ordine della serie storica iniziale di Prima
prima_ts_log <- log(prima_ts)
prima_ts_log_d1 <- diff(prima_ts_log, differences = 1)

# Grafico serie storica Prima logaritmica differenziata primo ordine
plot(prima_ts_log_d1, main = "Prima - Serie logaritmica Differenziata (1° Ordine)")

# Test ADF per Prima (log) differenziata primo ordine (superato se p-value < 0.05)
adf.test(prima_ts_log_d1) # p-value < 0.05 stazionaria

# La nuova serie storica di riferimento per Prima Assicurazioni
# è la serie logaritmica prima_ts_log
# Si passa nuovamente all'individuazione dei coefficienti e alla modellizzazione

# Analisi Autocorrelogrammi per Prima Logaritmica (stazionaria d=1)
par(mfrow=c(2,1))
acf(prima_ts_log_d1, main="ACF per Prima logaritmica differenziata (d=1)",
  lag.max=36)

```

```

pacf(prima_ts_log_d1, main="PACF per Prima logaritmica Differenziata (d=1)",
     lag.max=36)
par(mfrow=c(1,1))
# Possibile componente AR(2) o AR(3) in PACF; picco significativo ACF solo MA(3)
# Probabile ARIMA(2,1,0) o ARIMA(3,1,0) o al massimo ARIMA(2,1,1)

# Modellizzazione Prima Assicurazioni Logaritmica con auto.arima
modello_auto_log_prima <- auto.arima(
  prima_ts_log,
  seasonal = TRUE,
  stepwise = FALSE,
  approximation = FALSE
)
print(summary(modello_auto_log_prima)) # Output del modello
# Analisi Residui (Test Ljung-Box superato se p-value > 0.05)
checkresiduals(modello_auto_log_prima) # Diagnostica superata di poco
# Il modello stimato è un ARIMA(4,1,0) with Drift

# Modellizzazione "manuale" ARIMA(2,1,0)
modello_manual_log_prima1 <- Arima(prima_ts_log, order = c(2,1,0))
print(summary(modello_manual_log_prima1)) # Output del modello
# Analisi Residui (Test Ljung-Box superato se p-value > 0.05)
checkresiduals(modello_manual_log_prima1) # Diagnostica non superata

# Modellizzazione "manuale" ARIMA(2,1,1)
modello_manual_log_prima2 <- Arima(prima_ts_log, order = c(2,1,1))
print(summary(modello_manual_log_prima2)) # Output del modello
# Analisi Residui (Test Ljung-Box superato se p-value > 0.05)
checkresiduals(modello_manual_log_prima2) # Diagnostica non superata

# Modellizzazione "manuale" ARIMA(3,1,0)
modello_manual_log_prima3 <- Arima(prima_ts_log, order = c(3,1,0))
print(summary(modello_manual_log_prima3)) # Output del modello
# Analisi Residui (Test Ljung-Box superato se p-value > 0.05)
checkresiduals(modello_manual_log_prima3) # Diagnostica superata!

# 7.1 CONFRONTO MODELLI PRIMA -----

# Si creano due modelli A e B per il confronto
modello_prima_A <- modello_auto_log_prima # ARIMA(4,1,0) with Drift
modello_prima_B <- modello_manual_log_prima3 # ARIMA(3,1,0)

# Confronto modelli (Coeff e AIC e BIC, si preferiscono i valori più bassi)
print(summary(modello_prima_A))
print(summary(modello_prima_B))
# Il modello A presenta coefficienti AIC e BIC minori

# Confronto distribuzione dei residui rispetto a una Normale mediante QQ-Plot
par(mfrow=c(1,2))
# QQ plot Modello A
qqnorm(residuals(modello_prima_A), main="Q-Q Plot Residui Prima: Modello A")
qqline(residuals(modello_prima_A))
# QQ plot Modello B

```

```

qqnorm(residuals(modello_prima_B), main="Q-Q Plot Residui Prima: Modello B")
qqline(residuals(modello_prima_B))
par(mfrow=c(1,1))
# QQ-plot mostra una distribuzione nettamente migliore per il Modello A

# Confronto dati Out-of-Sample
# Teniamo gli ultimi 12 mesi come set di test
start_time_log <- start(prima_ts_log)
end_time_log <- end(prima_ts_log)
train_set_prima_log <- window(prima_ts_log, end = c(end_time_log[1],
                                                    end_time_log[2] - 12))
test_set_prima <- window(prima_ts, start = c(end_time_log[1],
                                             end_time_log[2] - 11))
# Notare che il train set è in scala log mentre il test set è sui dati originari

# Stima modelli sul solo training set
modello_prima_A_train <- Arima(train_set_prima_log, order=c(4,1,0),
                               include.drift=TRUE)
modello_prima_B_train <- Arima(train_set_prima_log, order=c(3,1,0))

# Generiamo le previsioni per i prossimi 12 mesi
forecast_prima_A <- forecast(modello_prima_A_train, h=12)
forecast_prima_B <- forecast(modello_prima_B_train, h=12)

# Riportiamo le previsioni alla scala originale usando la funzione exp()
forecast_prima_A$mean <- exp(forecast_prima_A$mean)
forecast_prima_A$lower <- exp(forecast_prima_A$lower)
forecast_prima_A$upper <- exp(forecast_prima_A$upper)
forecast_prima_A$x <- exp(forecast_prima_A$x)

forecast_prima_B$mean <- exp(forecast_prima_B$mean)
forecast_prima_B$lower <- exp(forecast_prima_B$lower)
forecast_prima_B$upper <- exp(forecast_prima_B$upper)
forecast_prima_B$x <- exp(forecast_prima_B$x)

# Confrontiamo l'accuratezza delle previsioni rispetto ai dati reali del test set
# La metrica più importante qui è RMSE (Root Mean Squared Error),
# si preferiscono i valori più bassi
print("Accuratezza Modello A (ARIMA (4,1,0) with drift su log)")
accuracy(forecast_prima_A, test_set_prima)

print("Accuratezza Modello B (Arima (3,1,0) su log)")
accuracy(forecast_prima_B, test_set_prima)
# La previsione è migliore nel modello A

# CONCLUSIONE CONFRONTO -
# Per Prima si sceglie il modello A : ARIMA(4,1,0) with Drift

# -----
# 8. STIMA DI IMPATTO DELLA CAMPAGNA PUBBLICITARIA DI PRIMA SU ALLIANZ (ARIMAX)
# -----

# Preparazione della variabile di intervento

```

```

# Convertiamo l'indice temporale numerico in un vettore di date reali
date_vector <- as.Date(as.yearmon(time(allianz_ts)))
# Definiamo la data del nostro intervento
date_of_intervention <- as.Date("2024-05-01")
# Creiamo la variabile dummy (0 prima, 1 dopo)
intervention_var <- ifelse(date_vector >= date_of_intervention, 1, 0)
# Convertiamo la variabile in un formato ts per darla in pasto ad Arima
intervention_ts <- ts(intervention_var, start=start(allianz_ts), frequency=12)

# Stima del modello di intervento

# Usiamo il nostro modello campione per Allianz e aggiungiamo l'intervento
# il modello è un SARIMA(1,1,1)(1,0,0)[12]
modello_intervento_allianz <- Arima(
  allianz_ts,
  order = c(1, 1, 1),
  seasonal = list(order = c(1, 0, 0), period = 12),
  xreg = intervention_ts
)

# Analisi del Risultato Finale
print("--- Risultato dell'Analisi di Intervento ---")
print(summary(modello_intervento_allianz)) # Il coefficiente di impatto è negativo

# Controlliamo anche i residui del modello finale
checkresiduals(modello_intervento_allianz) # La diagnostica è superata

# Test di significatività dell'impatto

# I nostri dati di input dall'output del modello
xreg_allianz <- -4.7770
xreg_allianz_se <- 5.0212
# Calcoliamo il t-statistic (o z-score, sono quasi identici per campioni grandi)
t_stat_allianz <- xreg_allianz / xreg_allianz_se
print(paste("Il t-statistic calcolato è:", round(t_stat_allianz, 4)))
# Calcoliamo il p-value
p_value <- 2 * pnorm(-abs(t_stat_allianz))
print(paste("Il p-value calcolato è:", round(p_value, 4)))
# p-value > 0.05 indica che il coefficiente non è statisticamente significativo

# Il modello ha rilevato un leggero calo nell'interesse di ricerca per Allianz
# nel periodo successivo al lancio della campagna.
# Tuttavia, questo calo non è statisticamente significativo.
# L'impatto osservato è indistinguibile da una normale fluttuazione casuale

# -----
# 9. STIMA DI AUTO-IMPATTO SU PRIMA ASSICURAZIONI
# -----

# Stima del modello di intervento

# Usiamo il nostro modello campione per Prima e aggiungiamo l'intervento
# Il modello è un ARIMA(4,1,0) with Drift

```

```

modello_intervento_prima <- Arima(
  prima_ts_log,
  order = c(4, 1, 0),
  include.drift = TRUE,
  xreg = intervention_ts
)

# Analisi del Risultato Finale
print("--- Risultato dell'Analisi di Intervento ---")
print(summary(modello_intervento_prima)) # Il coefficiente di impatto è negativo

# Test di significatività dell'impatto

# I nostri dati di input dall'output del modello
xreg_prima <- -0.3137
xreg_prima_se <- 0.1490
# Calcoliamo il t-statistic (o z-score, sono quasi identici per campioni grandi)
t_stat_prima <- xreg_prima / xreg_prima_se
print(paste("Il t-statistic calcolato è:", round(t_stat_prima, 4)))
# Calcoliamo il p-value
p_value_prima <- 2 * pnorm(-abs(t_stat_prima))
print(paste("Il p-value calcolato è:", round(p_value_prima, 4)))
# p-value < 0.05 indica che il coefficiente è statisticamente significativo

# Quantificazione stima dell'impatto
impact_prima <- (exp(xreg_prima) - 1)*100
print(paste("L'impatto è stato del:", round(impact_prima, 2), "%"))

# L'impatto della campagna è stato negativo e statisticamente significativo
# L'impatto è quantificato in -27% circa

# -----
# 10. ANALISI PREVISIONALE COMPARATIVA
# -----

# Preparazione della variabile di intervento futura
# Per prevedere, dobbiamo dire al modello quale sarà il valore di xreg nel futuro.
# Sappiamo che la campagna dura altri 5 mesi, poi cessa.
future_xreg <- c(rep(1, 5), rep(0, 7))

# Generiamo le previsioni dai modelli di intervento finali
forecast_allianz_finale <- forecast(modello_intervento_allianz, h = 12,
                                   xreg = future_xreg)
forecast_prima_log_finale <- forecast(modello_intervento_prima, h = 12,
                                     xreg = future_xreg)

# Trasformiamo la previsione di Prima in scala originale
forecast_prima_finale <- forecast_prima_log_finale
forecast_prima_finale$mean <- exp(forecast_prima_log_finale$mean)
forecast_prima_finale$lower <- exp(forecast_prima_log_finale$lower)
forecast_prima_finale$upper <- exp(forecast_prima_log_finale$upper)
forecast_prima_finale$x <- exp(prima_ts_log)

```



```

# Prepariamo i dati per il grafico con ggplot2

# Convertiamo la previsione in un semplice vettore numerico con as.numeric()
df_allianz_fc <- data.frame(Data=as.Date(as.yearmon(time(forecast_allianz_finale$mean))),
                             Previsione=as.numeric(forecast_allianz_finale$mean),
                             Lower=as.numeric(forecast_allianz_finale$lower[,2]),
                             Upper=as.numeric(forecast_allianz_finale$upper[,2]),
                             Brand="Allianz")

df_prima_fc <- data.frame(Data=as.Date(as.yearmon(time(forecast_prima_finale$mean))),
                           Previsione=as.numeric(forecast_prima_finale$mean),
                           Lower=as.numeric(forecast_prima_finale$lower[,2]),
                           Upper=as.numeric(forecast_prima_finale$upper[,2]),
                           Brand="Prima")

# Uniamo in un unico Dataframe previsioni per entrambi i brand
df_forecast <- rbind(df_allianz_fc, df_prima_fc)

# Dati storici per entrambi i brand
df_storico_allianz <- data.frame(Data=as.Date(time(allianz_ts)),
                                  Valore=as.numeric(allianz_ts), Brand="Allianz")
df_storico_prima <- data.frame(Data=as.Date(time(prima_ts)),
                                Valore=as.numeric(prima_ts), Brand="Prima")
df_storico <- rbind(df_storico_allianz, df_storico_prima)

# Creiamo il grafico finale
ggplot() +
  geom_ribbon(data=df_forecast, aes(x=Data, ymin=Lower, ymax=Upper, fill=Brand),
             alpha=0.2) +
  geom_line(data=df_storico, aes(x=Data, y=Valore, color=Brand),
            linetype="solid", linewidth=1) +
  geom_line(data=df_forecast, aes(x=Data, y=Previsione, color=Brand),
            linetype="dashed", linewidth=1.2) +
  geom_vline(xintercept = as.Date("2024-05-01"), linetype = "dotted",
             color = "red") +
  annotate("text", x = as.Date("2024-05-01"), y = 50,
           label = "Lancio Campagna Prima (6 Mag '24)", color="red", angle=90,
           vjust = -0.5, size=3) +
  labs(
    title="Analisi Storica e Previsionale con Intervalli di Confidenza al 95%",
    subtitle="Previsioni condizionali basate sull'impatto e la durata della campagna marketing.",
    x="Data",
    y="Indice di Interesse (Google Trends)",
    caption = "Analisi basata su modelli ARIMAX di intervento."
  ) +
  scale_color_manual(values = c("Allianz" = "blue", "Prima" = "green"),
                     aesthetics = c("color", "fill")) +
  theme_light()

# =====
# --- FINE ---
# =====

```