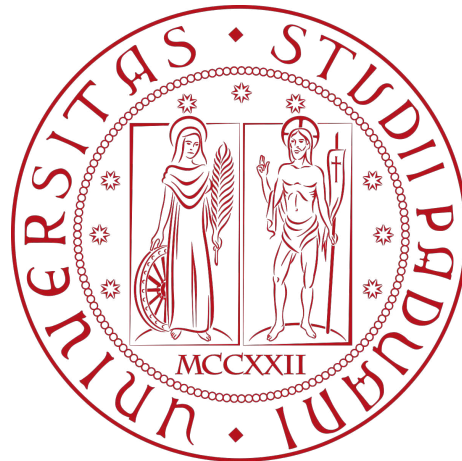


Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica per l'Economia e l'Impresa



**Previsione aziendale: il caso di un'attività commerciale nel
settore fai da te**

Relatrice Prof.ssa Mariangela Guidolin
Dipartimento di Scienze Statistiche

Laureando: Michele Carini
Matricola N° 2042559

Anno Accademico 2023/2024

Indice

Introduzione	1
1 Processo di previsione aziendale	5
1.1 Definizione degli obiettivi	5
1.2 Raccolta delle informazioni	6
1.3 Analisi esplorative	9
1.3.1 Visualizzazione grafica	9
1.3.2 Autocorrelazione	10
1.4 Scelta e stima del modello	11
1.5 Valutazione del modello	14
1.5.1 Controllo diagnostico dei residui	14
1.5.2 Criteri per il confronto tra modelli	15
1.5.3 Misure di accuratezza delle previsioni	15
2 Gestione della stagionalità nella domanda	17
2.1 Caso studio: pennelli	18
2.1.1 Serie trimestrale	18
2.1.2 Serie mensile	24
3 Gestione del rischio di obsolescenza	37
3.1 Caso studio: duplicati di chiavi	37
4 Analisi delle tendenze di mercato	45
4.1 Caso studio: smalti per legno e metallo	46
4.1.1 Scenari di previsione	49

5	Conclusioni	55
A	Codice R	57
A.1	Modelli del capitolo 1	57
A.2	Modelli del capitolo 2	60
A.3	Modelli del capitolo 3	61
B	Modelli di previsione	63
B.1	Modello di regressione per la stima di trend e stagionalità	63
B.2	Modelli di lisciamiento esponenziale	64
B.2.1	Modello di Holt con trend smorzato	64
B.2.2	Modello di Holt-Winters con trend smorzato e stagionalità moltiplicativa	65
B.3	Modelli <i>SARIMA</i>	65
B.4	Modello di regressione dinamica SARIMAX	67
B.5	Modelli di diffusione	68
B.5.1	Modello di Bass standard	68
B.5.2	Modello di Bass generalizzato con shock esponenziale	69
C	Test statistici e criteri per il confronto tra modelli	71
C.1	Test per l'incorrelazione dei residui	71
C.1.1	Test di Durbin-Watson	71
C.1.2	Test di Ljung-Box	73
C.2	Criteri per la selezione del modello	73
C.2.1	R^2 e R^2 corretto	73
C.2.2	Criteri di informazione	74
	Bibliografia	74
	Sitografia	76

Introduzione

Questa relazione finale unisce approcci statistici ed esempi pratici. Alcuni dei modelli di previsione più comuni verranno, infatti, collaudati e testati sulle serie storiche delle vendite di un'attività commerciale del settore del Fai Da Te. In particolare, la ditta Carini S.R.L., nata a Sanremo negli anni '60, è una ferramenta specializzata nel colore per l'edilizia e la carrozzeria. L'attività possiede una rete di agenti che vende e consegna le merci sul territorio ligure e in Francia. Come tante imprese del settore, l'azienda opera la vendita sia al dettaglio che all'ingrosso. Questo aspetto risulta determinante nelle decisioni aziendali, le quali devono andare incontro ai bisogni sia dell'acquirente comune sia dei professionisti. L'obiettivo di questo elaborato è quello di scoprire quali siano i modelli statistici di previsione più efficaci nel contesto in esame, in modo tale da ottimizzare i futuri processi di acquisto e la gestione del magazzino. Il mercato del Fai Da Te coinvolge, in Italia, un numero molto elevato di imprese, artigiani, rivenditori e acquirenti. In figura 1 è rappresentato l'andamento dal 2013 al 2021 del commercio all'ingrosso e al dettaglio di vernici, colori e ferramenta. Il dato più positivo riguarda certamente la vendita al dettaglio, che è in continua crescita negli ultimi anni; la vendita all'ingrosso, sebbene abbia riscontrato un calo dovuto alla pandemia di Covid19, mantiene una quota di fatturato elevata. Nonostante la serie Istat si fermi al 2021, si vede chiaramente come proprio nell'ultimo anno registrato ci sia stata una ripresa economica dell'intero settore. Una ricerca di Assofermet e CreditSafe (Assofermet 2023) ha riportato che tra il 2019 e il 2022 il fatturato delle imprese del settore ferramenta, Fai Da Te e bricolage è cresciuto del 34,6%.

Un ulteriore dato a conferma della recente crescita del settore è l'indice Istat delle vendite al dettaglio per il gruppo 'utensileria per la casa e ferramenta', il cui andamento è rappresentato in figura 2.

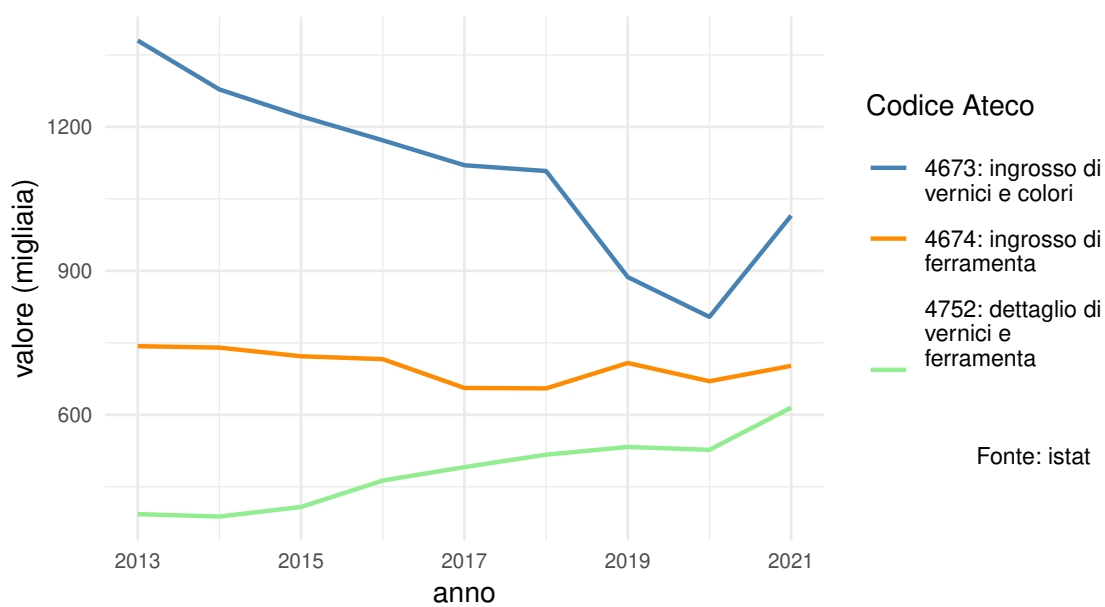
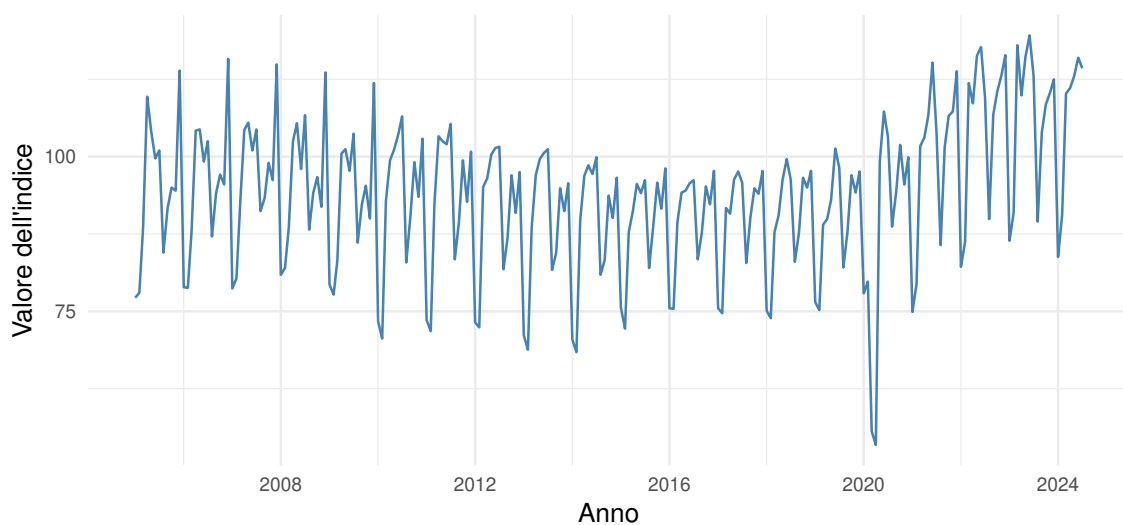


Figura 1: Risultati economici annuali delle imprese con classe di addetti 10-19 dal 2013 al 2021 per i codici Ateco 2007 riportati in legenda.

Il trend positivo degli ultimi anni ha interessato anche l'azienda oggetto di questo studio (figura 3). Infatti, i dati a disposizione mostrano un notevole incremento dei ricavi, in particolare per quanto riguarda l'ingrosso. Inoltre negli ultimi anni si è creata una netta differenza tra i due canali di distribuzione; eccetto i mesi di marzo e aprile 2020, i più difficili per imprese e artigiani in tempi di pandemia, i valori mensili del fatturato all'ingrosso sono circa tre volte quelli dei ricavi derivanti dalla vendita al dettaglio.



Fonte: Istat

Figura 2: andamento dell'indice Istat relativo alle vendite al dettaglio del gruppo 'utensileria per la casa e ferramenta'; dati mensili con base 2021.

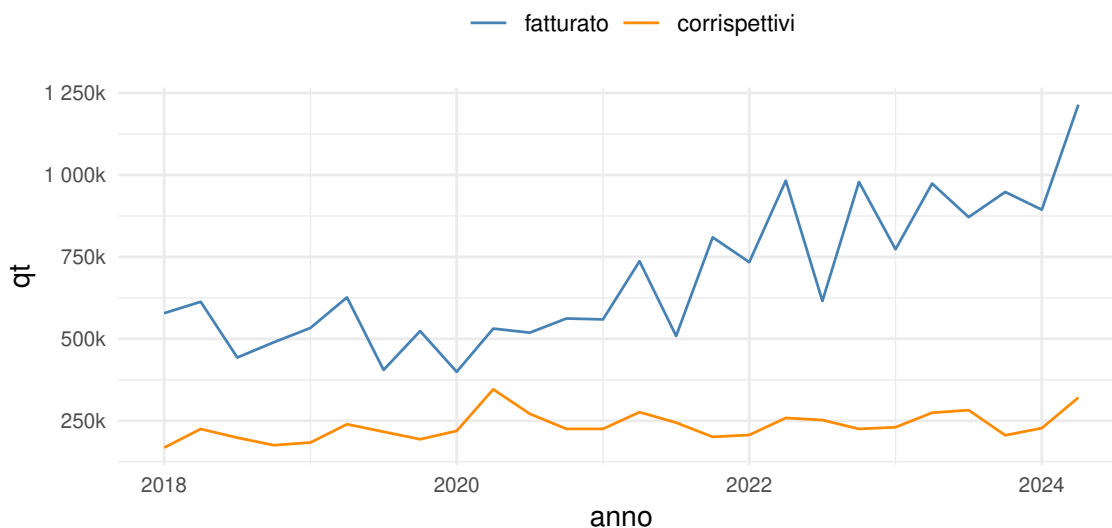


Figura 3: grafico dei ricavi trimestrali divisi per fatturato e corrispettivi da gennaio 2018 ad agosto 2024 dell'azienda Carini S.R.L.

A partire da queste prime considerazioni è nata l'idea di analizzare i dati a disposizione attraverso modelli di previsione adeguati. La struttura della relazione è così definita:

- Capitolo 1: in questo capitolo verranno delineati tutti i passaggi di un processo di previsione in un contesto aziendale. Le tecniche descritte sono quelle che saranno utilizzate per l'analisi delle principali problematiche aziendali attraverso i casi studio dei capitoli successivi.
- Capitolo 2: verrà approfondito il tema della variabilità stagionale nella domanda. In questo capitolo saranno proposti diversi modelli per la previsione delle vendite di pennelli, come il modello di regressione lineare multipla, il modello di Holt-Winters e il modello SARIMAX.
- Capitolo 3: il tema di questo capitolo è il rischio di obsolescenza dei prodotti. Alla serie delle vendite di chiavi a profilo universale verranno adattati due modelli di diffusione in grado di prevedere adeguatamente le fasi finali del un ciclo di vita di un prodotto. In particolare, verrà stimato un modello di Bass Standard con errori SARIMA e un modello di Bass generalizzato con shock esponenziale.
- Capitolo 4: in questo capitolo l'attenzione sarà posta sulle tendenze del mercato e sui fattori che le determinano. Il tema sarà trattato attraverso l'analisi di due scenari di previsione per una serie storica delle vendite di vernici all'acqua e a solvente.

Capitolo 1

Processo di previsione aziendale

In un contesto aziendale, un processo di previsione statisticamente accurato è fondamentale per supportare la pianificazione strategica e ottimizzare l'allocazione delle risorse. Di seguito sono illustrate le fasi principali di un processo di previsione aziendale (Hyndman, R.J. e Athanasopoulos, G. 2018):

1. definizione degli obiettivi
2. raccolta delle informazioni
3. analisi esplorative
4. scelta e stima del modello
5. valutazione del modello

1.1 Definizione degli obiettivi

Prima di tutto bisogna definire qual è l'obiettivo dell'analisi previsiva e di conseguenza il tipo di previsione che si vuole effettuare; le previsioni possono essere di breve, medio o lungo termine. In azienda solitamente

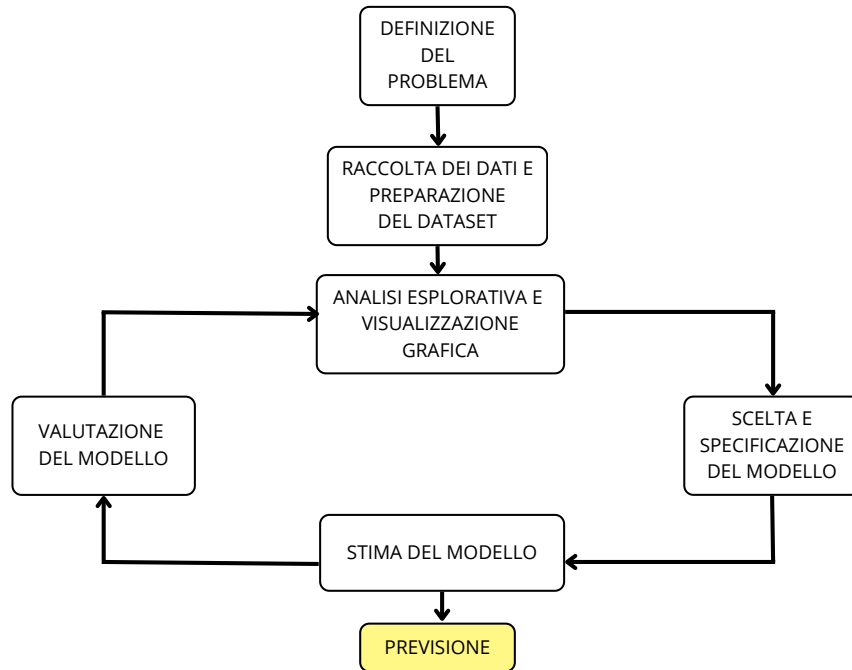


Figura 1.1: processo di previsione aziendale.

l'attenzione è posta principalmente su analisi di breve periodo; le serie storiche delle vendite, infatti, sono spesso molto brevi e caratterizzate dalla presenza di andamenti irregolari, influenzati da variabili di contesto non sempre identificabili. Questi aspetti non permettono solitamente di ottenere previsioni a medio e lungo termine sufficientemente affidabili.

1.2 Raccolta delle informazioni

In secondo luogo occorre estrarre i dati oggetto d'interesse. Ormai in tutte le attività commerciali i dati interni sono disponibili nel software gestionale. Le serie storiche che verranno utilizzate nei capitoli successivi sono state esportate in file excel proprio dal gestionale

dell'azienda. Questa fase è fondamentale perchè permette a chi deve svolgere le analisi statistiche di scegliere le variabili e il periodo di tempo da considerare in relazione al tipo di previsione da formulare. Una volta caricati i dati in un software statistico, è necessario trasformarli per poter condurre le analisi richieste; in questo caso sono stati utilizzati i pacchetti di R *dplyr* e *tidyr*. Di seguito, si propone un esempio di codice R utilizzato per la manipolazione dei dati di vendita di bobine di carta industriali:

```
library(readxl)
library(dplyr)
library(tidyr)

# carico i dati delle vendite dal file excel
carta <- read_excel('bobina_carta.xlsx')
head(carta)
attach(carta)
# A tibble: 6 x 4
  data          tipo_doc          qt valore
<dtm>         <chr>          <dbl> <dbl>
1 2005-01-03    FATTURA          2  16.7
2 2005-01-04    SCONTRINO FISCALE  1   8.33
3 2005-01-05    SCONTRINO FISCALE  2  16.7
4 2005-01-17    FATTURA          2  16.7
5 2005-01-24    SCONTRINO FISCALE  2  16.7
6 2005-01-27    SCONTRINO FISCALE  1   8.33

# Converto la colonna della data in formato Date
carta$data <- as.Date(data)
# estraggo l'anno e il trimestre di ogni data
carta$anno <- year(data)
carta$trim <- quarter(data)

# Raggruppo per tipo documento e trimestre, calcolando
# la somma degli scarichi (qt)
carta_agg <- carta %>%
  group_by(anno, trim, tipo_doc) %>%
  summarise(tot_scarichi = sum(qt, na.rm = TRUE))
head(carta_agg)
# A tibble: 6 x 4
# Groups:   anno, trim [3]
```

```

  anno  trim  tipo_doc          tot_scarichi
  <dbl> <int> <chr>                <dbl>
1  2005     1 FATTURA             28
2  2005     1 SCONTRINO FISCALE       17
3  2005     2 FATTURA             17
4  2005     2 SCONTRINO FISCALE       13
5  2005     3 FATTURA             28
6  2005     3 SCONTRINO FISCALE       18

# trasformo le categorie del parametro
# 'tipo_documento' in due variabili differenti
carta.finale <- carta_agg %>%
  pivot_wider(names_from = tipo_doc, values_from = tot_
    scarichi, values_fill = list(tot_scarichi = 0))
colnames(carta.new) <- c('anno', 'trimestre',
  'fattura', 'scontrino')

head(carta.finale)
# A tibble: 6 x 4
# Groups:   anno, trimestre [6]
  anno trimestre fattura scontrino
  <dbl>     <int>   <dbl>     <dbl>
1  2005         1     28         17
2  2005         2     17         13
3  2005         3     28         18
4  2005         4     19         10
5  2006         1     29         27
6  2006         2     26         12

# trasformo i dati in un oggetto di tipo serie storica
carta.ts <- ts(carta.new[1:78, 3:4], start = c(2005, 1),
  frequency = 4)

```

Oltre ai dati interni dell'azienda, in questa fase può essere necessario raccogliere dati esterni, come quelli socio-economici o climatici. Questo tipo di informazioni, infatti, possono influire in modo significativo sulle aziende, modificando, per esempio, la domanda di acquisto di qualche bene o servizio.

1.3 Analisi esplorative

1.3.1 Visualizzazione grafica

Dopo aver definito gli obiettivi e raccolto le informazioni necessarie, il passo successivo consiste nel visualizzare graficamente la serie storica.

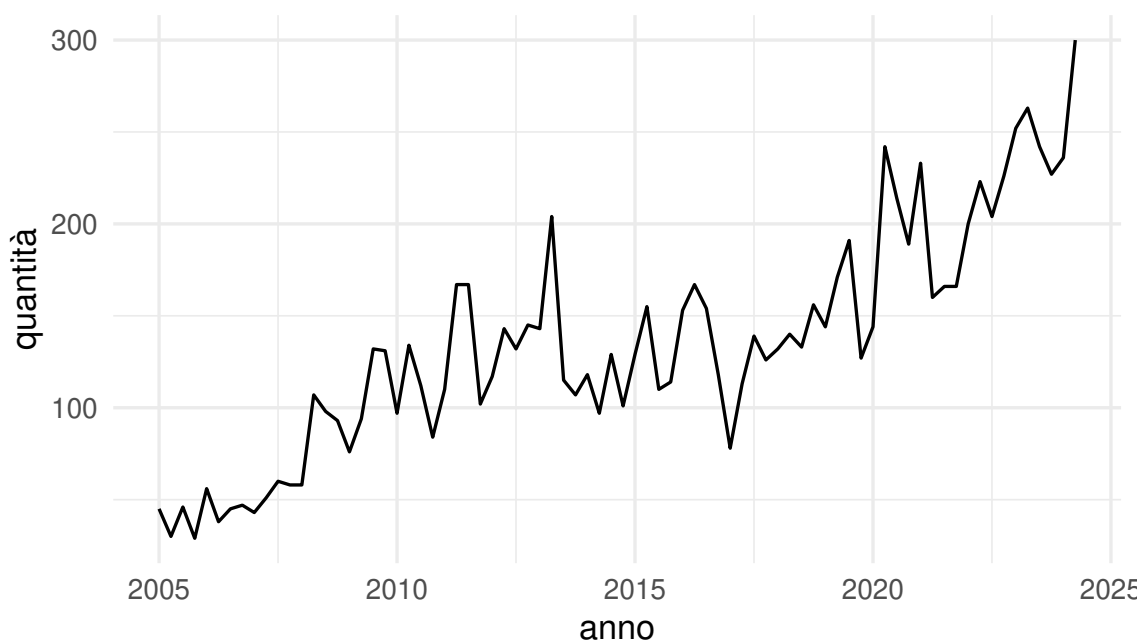


Figura 1.2: serie storica delle vendite trimestrali di bobine di carta industriali dal 2005 al 2024.

Già dal grafico delle serie si possono fare numerose considerazioni. Per esempio, dal grafico in figura 1.1, è possibile vedere la presenza di un trend positivo che si estende con un'intensità variabile per tutta la serie. Inoltre, i valori non sembrano seguire un particolare schema stagionale. Per indagare meglio sul comportamento della serie può essere utile anche visualizzare l'andamento in relazione ad altre variabili, come il tipo di documento.

Dal grafico in figura 1.3, per esempio, è possibile vedere come in realtà il volume delle vendite sia composto principalmente dal fatturato, la cui

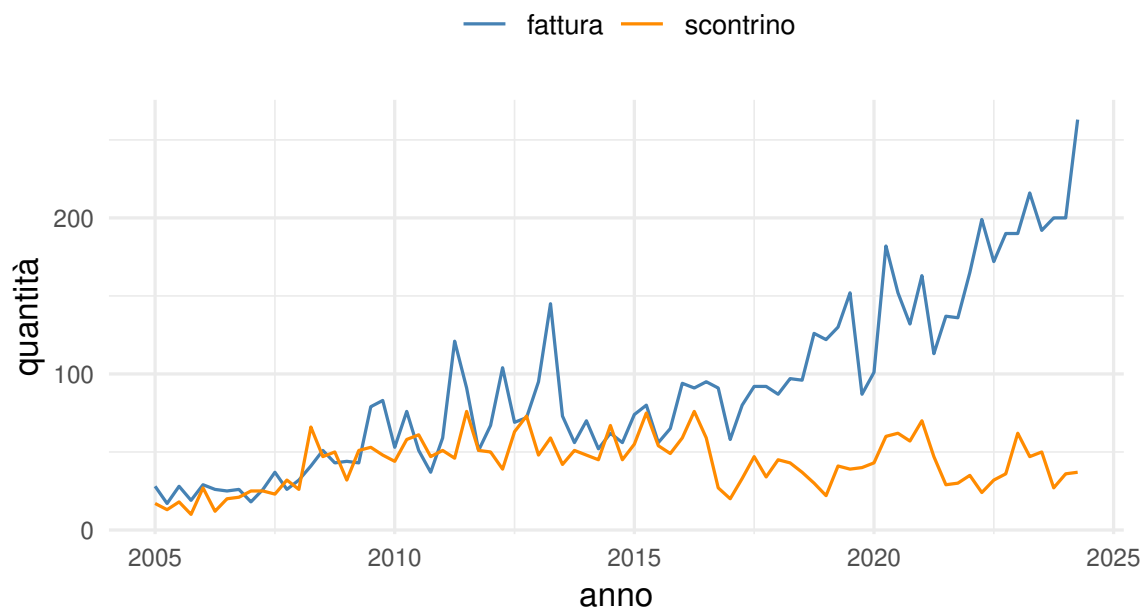


Figura 1.3: serie storica delle vendite di bobine di carta industriali suddivisa per tipo di documento.

serie sembra essere caratterizzata da un trend quasi esponenziale. La serie dei corrispettivi, al contrario, appare sostanzialmente piatta. Nei capitoli successivi verranno illustrate anche diverse tipologie di grafici per le serie interessate dalla componente stagionale.

1.3.2 Autocorrelazione

L'autocorrelazione misura la relazione lineare tra valori ritardati di una serie storica. Il coefficiente di autocorrelazione può essere scritto come:

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}, \quad (1.1)$$

dove k rappresenta il ritardo preso in considerazione. L'insieme dei coefficienti di una serie storica costituisce la funzione di

autocorrelazione (ACF). L'ACF risulta molto utile in fase preliminare per individuare le componenti di una serie storiche. Nel correlogramma in figura 1.4 i valori del coefficiente di autocorrelazione decrescono lentamente e hanno tutti un valore piuttosto elevato; questo indicherebbe la presenza di un trend significativo.

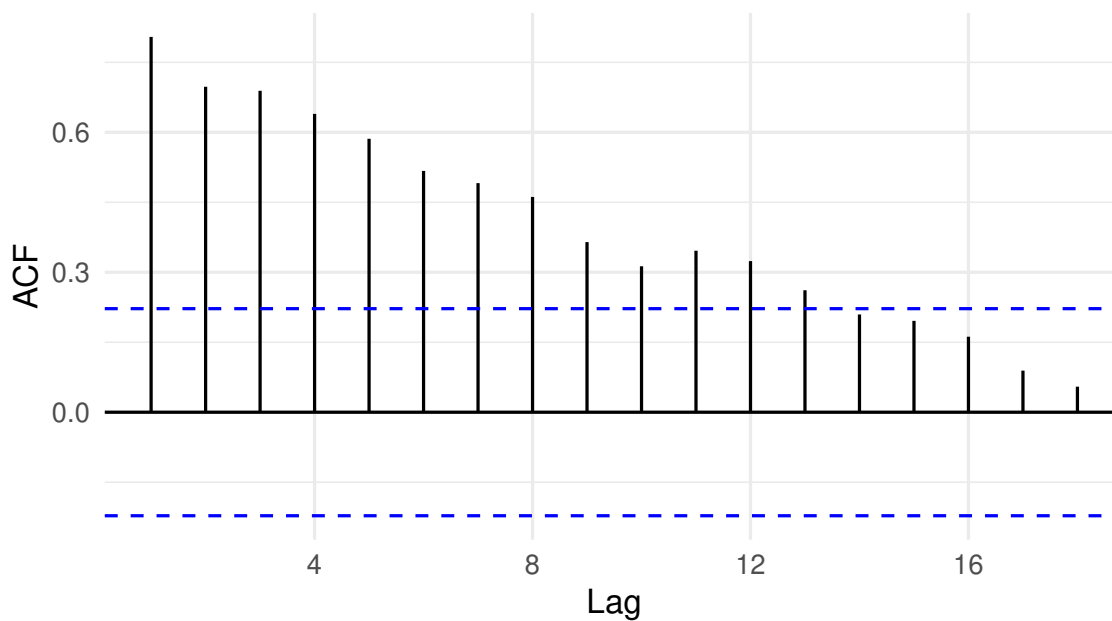


Figura 1.4: funzione di autocorrelazione per le vendite di bobine di carta.

Le linee blu tratteggiate rappresentano gli intervalli di confidenza al 95% e servono per stabilire se le correlazioni sono significativamente diverse da zero. In generale le serie di vendite tendono ad essere autocorrelate positivamente.

1.4 Scelta e stima del modello

Dopo le analisi esplorative occorre scegliere e successivamente stimare il modello che si ritiene più adatto per la problematica da affrontare, anche in relazione al tipo di dati storici che si hanno a disposizione. In

realtà, nella maggior parte dei casi è consigliabile confrontare almeno un paio di potenziali modelli prima di una scelta definitiva. In questa relazione verranno trattati i seguenti modelli:

- **Modelli di regressione lineare per serie storiche.**

I modelli di regressione si basano sull'assunzione che la variabile da prevedere sia legata da una relazione con una o più variabili indipendenti, dette predittori. Nel contesto delle serie storiche si possono includere come variabili esplicative il trend, la stagionalità e altre variabili esterne per spiegare l'andamento della variabile di interesse. Come modello esplicativo può risultare utile per analizzare e prevedere i cambiamenti nel tempo, ma potrebbe non catturare adeguatamente la dipendenza temporale come fanno i modelli della classe ARIMA.

- **Modelli di lisciamiento esponenziale.**

I modelli di lisciamiento sono semplici ed allo stesso tempo efficaci, soprattutto per analisi di previsione di breve periodo. Le previsioni costruite con i modelli di lisciamiento esponenziale sono medie pesate delle osservazioni passate, ottenute assegnando pesi maggiori alle osservazioni più recenti. I metodi di lisciamiento trattati saranno rappresentati nella forma per componenti, per cui avremo una funzione di previsione e un'equazione di lisciamiento per ogni componente (livello, trend, stagionalità), a seconda del modello.

- **Modelli della classe ARIMA e SARIMA.**

Nei modelli ARIMA la previsione del futuro si basa solamente sui valori storici della serie stessa e sugli errori passati, senza considerare variabili esplicative. Questa tipologia di modelli si adatta bene per serie stazionarie o che possono essere rese stazionarie attraverso differenziazione. In generale i modelli della classe ARIMA contengono i seguenti parametri:

- p (autoregressione): Indica il numero di termini autoregressivi, ossia quante osservazioni passate influenzano il valore attuale della serie.
- d (differenziazione): Indica il numero di differenziazioni necessarie per rendere la serie stazionaria, eliminando eventuali trend.
- q (media mobile): Indica il numero di termini di errore passati inclusi nel modello.

I modelli SARIMA sono un'estensione dei modelli ARIMA che consentono di introdurre parametri stagionali (P, D, Q).

- **Modelli di regressione dinamici.**

I modelli dinamici integrano una regressione con un modello ARIMA o SARIMA applicato agli errori, al fine di catturare l'autocorrelazione presente nei dati. Questo approccio migliora la precisione delle previsioni considerando sia fattori esogeni sia la struttura temporale degli errori.

- **Modelli di diffusione.**

I modelli di diffusione sono utilizzati per descrivere e prevedere il processo con cui nuovi prodotti, innovazioni o tecnologie si diffondono all'interno di un mercato o di una popolazione. L'obiettivo di questi modelli è quello di prevedere quando verrà raggiunto il picco di adozione e quanto tempo ci vorrà affinché la maggior parte della popolazione adotti il prodotto prima che la crescita rallenti. I modelli di diffusione, come il modello di Bass, distinguono i consumatori in innovatori, ovvero i primi acquirenti di un prodotto o servizio, e imitatori, ovvero la maggioranza della popolazione influenzata dal comportamento degli altri acquirenti.

Per un approfondimento dei modelli utilizzati nei casi studio si veda l'appendice B.

1.5 Valutazione del modello

Una volta stimato un modello, occorre verificare il suo adattamento rispetto ai dati. Esistono diversi strumenti diagnostici per la verifica del comportamento del modello, oltre a misure di accuratezza che permettono di confrontare tra loro più modelli.

1.5.1 Controllo diagnostico dei residui

In seguito alla stima del modello, è consuetudine analizzare i residui, ovvero la differenza tra le osservazioni e i valori stimati $e_t = y_t - \hat{y}_t$. In un modello di previsione statistico è necessario che i residui rispettino le seguenti caratteristiche:

- incorrelazione: se vi è ancora correlazione nei residui significa che il modello non ha catturato al meglio la struttura dei dati. Residui privi di autocorrelazione vengono detti white noise (rumore bianco). L'incorrelazione può essere testata facilmente attraverso un'analisi del correlogramma; se nel grafico dell'ACF ci sono più correlazioni significative per lag diversi da zero, allora il modello probabilmente può essere migliorato. Oltre all'analisi grafica è possibile effettuare alcuni test, come il test di Durbin-Watson o il test di Ljung-Box (si veda la sezione C.1 dell'appendice C).
- media nulla: se i residui del modello hanno media diversa da zero, le previsioni saranno distorte.
- omoschedasticità: residui con varianza costante rendono più affidabili le previsioni e il calcolo degli intervalli di confidenza.
- normalità: i residui devono avere distribuzione più possibile vicina ad una normale. Si può confrontare la distribuzione dei residui con una normale attraverso il Q-Q plot oppure si può condurre un test statistico come quello di Shapiro-Wilk o Jarque-Bera.

1.5.2 Criteri per il confronto tra modelli

Oltre ad elementi qualitativi, come la semplicità e l'interpretabilità, è necessario utilizzare dei criteri quantitativi per valutare e confrontare i modelli. I principali sono:

- Criteri di Informazione; sono utilizzati per confrontare modelli con diversa complessità. Per questi criteri, un valore più basso indica un modello migliore. I più utilizzati sono l'AIC (criterio di Informazione di Akaike) e il BIC (Criterio di Informazione Bayesiano).
- Coefficiente di determinazione R^2 per modelli esplicativi; è utile nei modelli di regressione per confrontare la bontà di adattamento di più modelli. L' R^2 corretto penalizza i modelli con tanti parametri (si veda la sezione C.2 dell'appendice C).

1.5.3 Misure di accuratezza delle previsioni

Per valutare in modo efficace l'accuratezza delle previsioni è importante realizzare un confronto tra le stime future del modello e i dati reali non utilizzati per l'adattamento, poichè la dimensione dei residui non è un'indicazione affidabile di quanto è probabile che siano grandi i veri errori di previsione. Quando si scelgono i modelli è infatti pratica comune separare i dati disponibili in due parti, un campione di stima (training set) e un campione test (test set), dove i dati di addestramento sono usati per stimare i parametri del metodo di previsione e i dati di test sono usati per valutarne la precisione. Idealmente, l'ampiezza del campione di test dovrebbe essere grande almeno quanto l'orizzonte di previsione richiesto. A partire dall'errore di previsione, definito come:

$$e_t = Y_t - F_t, \quad (1.2)$$

esistono diversi indicatori di accuratezza che possono essere usati sia sull'intero campione di dati a disposizione, sia sui campioni di stima e test. Le misure più utilizzate sono:

- Errore medio: solitamente è piccolo, dal momento che errori positivi e negativi si compensano; ci informa sulla presenza di distorsione nelle stime.

$$\text{ME} = \frac{1}{n} \sum_{t=1}^n e_t$$

- Radice quadrata degli errori di previsione al quadrato: viene utilizzato per l'ottimizzazione dei parametri.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2}$$

- Errore medio assoluto: è la misura più interpretabile e più utilizzata per il confronto tra modelli.

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |e_t|$$

Per il confronto di metodi di previsione differenti viene spesso preso in considerazione l'errore percentuale, dato da $pe_t = 100e_t/Y_t$ e le derivanti misure di accuratezza percentuali:

- Errore medio percentuale:

$$\text{MPE} = \frac{1}{n} \sum_{t=1}^n pe_t$$

- Errore medio assoluto percentuale:

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n |pe_t|.$$

Capitolo 2

Gestione della stagionalità nella domanda

Quasi tutte le attività commerciali devono fronteggiare il problema della stagionalità. In realtà, spesso, le variazioni della domanda dovute a fattori stagionali possono essere prevedibili. Nel contesto della gestione del magazzino e delle scorte, riconoscere e gestire efficacemente la stagionalità è cruciale per evitare rotture di stock durante i picchi di domanda e accumuli di magazzino durante i periodi in cui la domanda è più contenuta o nulla. L'ottimizzazione delle scorte rappresenta uno degli aspetti fondamentali per il successo operativo e commerciale di un'azienda. Una buona gestione delle scorte, infatti, consente di rispondere alla domanda dei clienti e garantire la continuità delle operazioni di vendita. Al contrario, una gestione inefficace del magazzino può comportare costi significativi, inclusi costi di stoccaggio e costi legati al rischio di deterioramento della merce, con un impatto negativo sulla redditività aziendale. In questo capitolo verranno utilizzati modelli statistici in grado di riconoscere i pattern stagionali tipici delle serie storiche delle vendite mensili e trimestrali.

2.1 Caso studio: pennelli

In questo caso studio verrà affrontato il tema della stagionalità nella domanda attraverso la serie storica trimestrale e poi mensile delle vendite di pennelli. I dati utilizzati sono di tipo aggregato, in quanto inglobano le principali tipologie di pennelli venduti: pennelli piatti, angolari, tondi e pennelli da radiatore.

2.1.1 Serie trimestrale

L'obiettivo di questa prima analisi è quello di descrivere e prevedere l'andamento generale della serie trimestrale delle vendite di pennelli. Dato che le osservazioni a disposizione sono relativamente poche, è stato scelto di adattare un modello di regressione lineare multipla.

Analisi esplorativa

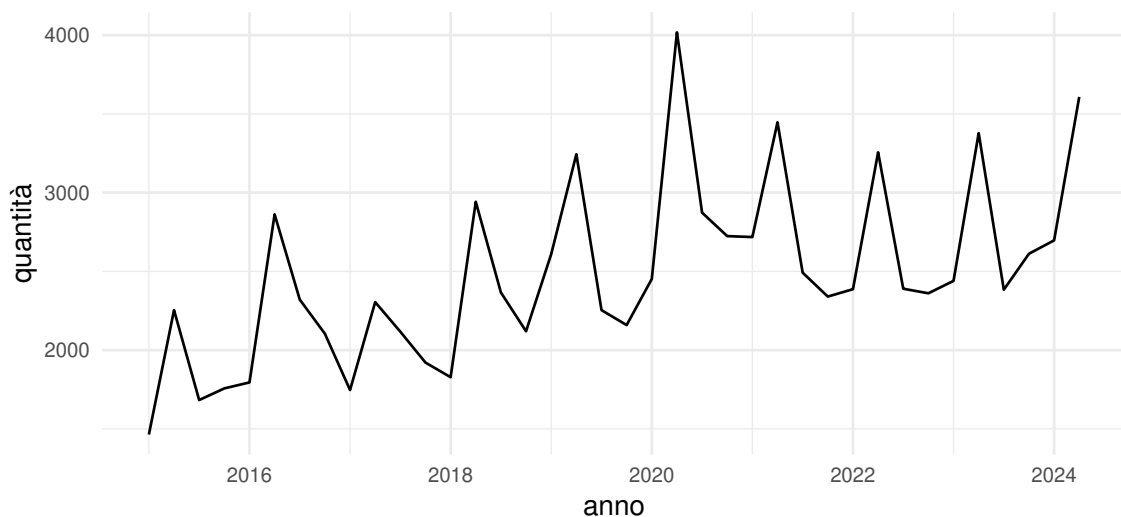


Figura 2.1: serie delle vendite trimestrali di pennelli dal 2015 al secondo trimestre del 2024.

Osservando il grafico in figura 2.1 possiamo già fare alcune considerazioni. La serie è certamente caratterizzata da una stagionalità piuttosto regolare. Inoltre il grafico evidenzia un trend crescente e apparentemente lineare. Infine, è chiara la presenza di alcuni valori nell'anno 2020 e 2021 che si discostano significativamente in modo positivo dall'andamento generale delle vendite. Il periodo in questione è quello del lockdown dovuto alla pandemia di Covid19, in cui i negozi di ferramenta sono stati autorizzati all'apertura anche nei mesi più critici. Molti consumatori privati hanno approfittato del tempo libero per effettuare lavori di manutenzione domestica e riparazioni. In alcuni casi studio successivi si vedrà come gli effetti della quarantena abbiano, invece, provocato un decremento delle vendite.

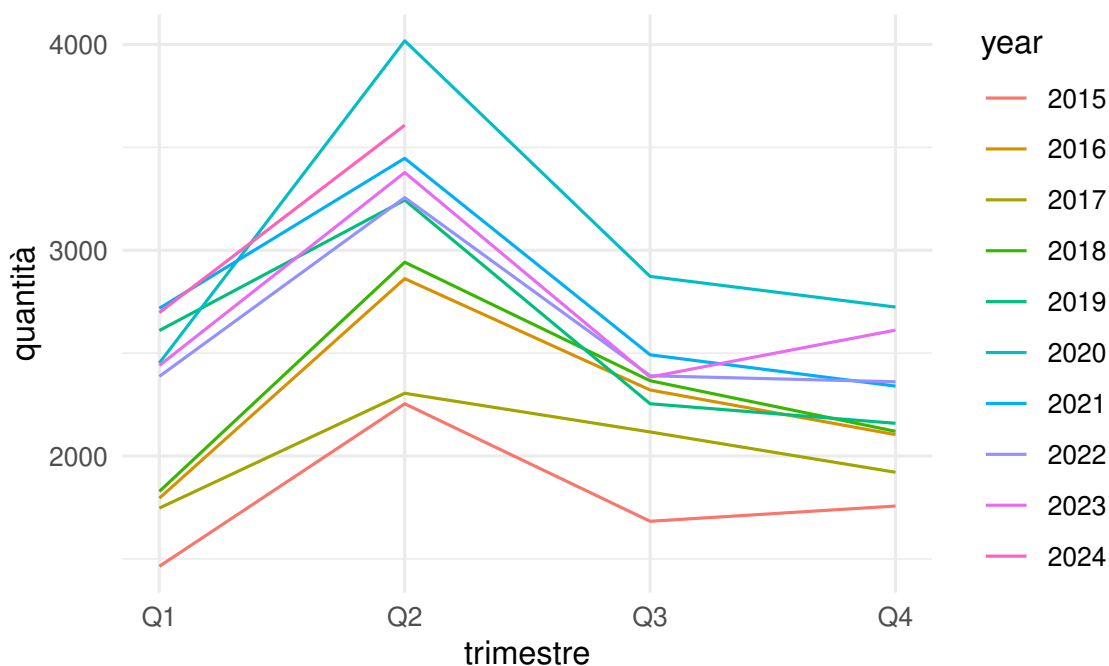


Figura 2.2: vendite di pennelli suddivise per trimestre.

Il grafico in figura 2.2 ci permette di individuare in maniera più precisa qual è il comportamento della componente stagionale all'interno della serie. Risulta subito molto evidente come il secondo trimestre abbia

valori tendenzialmente molto più alti degli altri tre. Anche qui si può notare come il 2020, in particolare negli ultimi tre trimestri, sia caratterizzato dalla presenza di valori ben oltre la media stagionale.

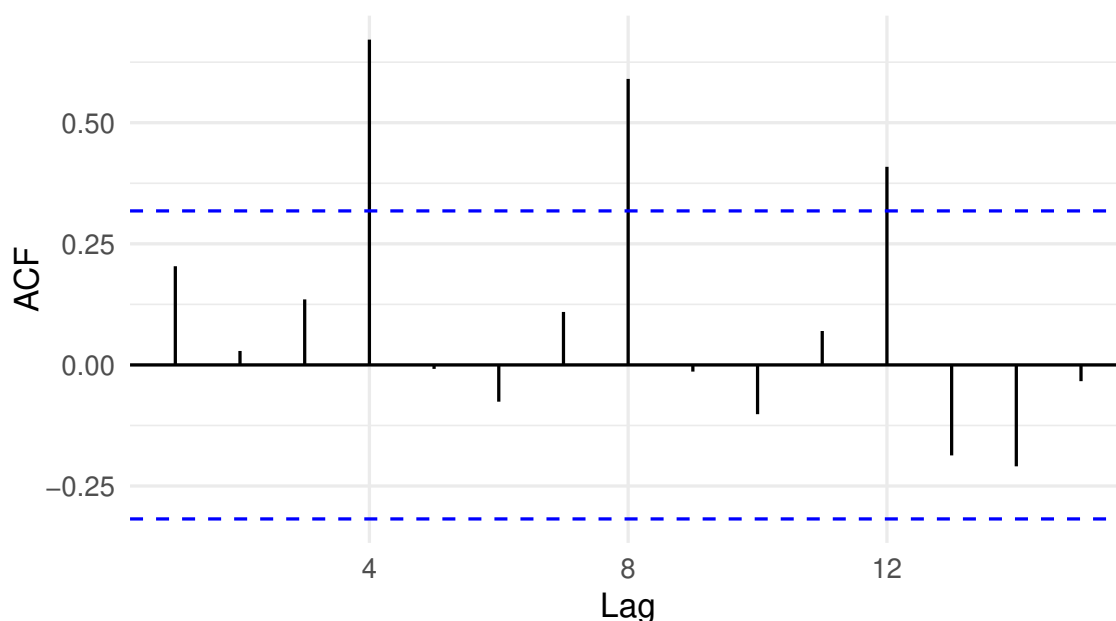


Figura 2.3: funzione di autocorrelazione per la serie delle vendite trimestrali di pennelli.

La funzione di autocorrelazione (figura 2.3) evidenzia la presenza di picchi positivi significativi in corrispondenza dei ritardi stagionali, confermando la presenza di periodicità nella serie.

Modello di regressione con variabili dummy

Nei modelli di regressione lineare la stagionalità viene solitamente trattata attraverso l'impiego di variabili dummy. È importante sottolineare che, per evitare il problema della perfetta multicollinearità, è necessario includere nel modello $s - 1$ variabili dummy stagionali, dove s costituisce il periodo della serie. Nel caso in cui il processo generatore della serie sia caratterizzato simultaneamente da trend e stagionalità, si avrà:

$$Y_t = f(t) + g(t) + \epsilon_t,$$

con $f(t)$ funzione che descrive il trend e $g(t)$ funzione che rappresenta la componente stagionale. Dato che la serie oggetto di studio è trimestrale ($s = 4$), il modello avrà la seguente forma:

$$y_t = \beta_0 + \beta_1 t + \beta_2 d_{2t} + \beta_3 d_{3t} + \beta_4 d_{4t} + \beta_5 d_{ct} + \epsilon_t, \quad (2.1)$$

dove l'ultima variabile dummy d_{ct} assume valore 1 nei trimestri influenzati dalla pandemia. I coefficienti $\beta_2, \beta_3, \dots, \beta_s$ misurano la variazione rispetto a β_0 imputabile alla stagionalità al netto dell'effetto degli altri coefficienti inclusi nel modello. Le stime dei coefficienti del modello, rappresentate nella tabella 2.1, sono state ottenute con il metodo dei minimi quadrati lineari.

coeff	stima	errore st.	statistica t	p-value
β_0	1661.937	92.029	18.059	$< 2 \times 10^{-16}$
β_1	26.078	3.242	8.043	3.5×10^{-9}
β_2	891.522	97.201	9.172	1.8×10^{-10}
β_3	99.935	99.818	1.001	0.324
β_4	-13.032	99.865	-0.130	0.897
β_5	563.883	115.756	4.871	2.8×10^{-5}

Tabella 2.1: stime dei coefficienti del modello 2.1.

Nonostante siano poco significativi dal punto di vista statistico, gli ultimi due coefficienti stagionali contribuiscono a rendere completo il modello, evitando di trascurare comportamenti stagionali potenzialmente importanti.

Le statistiche del modello in tabella 2.2 riportano nel complesso valori buoni, eccetto la statistica di Durbin-Watson, la quale rileva presenza di autocorrelazione seriale positiva di ordine uno nei residui del modello. Problemi di autocorrelazione sono confermati anche dal grafico e dalla funzione di autocorrelazione dei residui in figura 2.5. Inoltre, sempre dallo stesso grafico è possibile identificare la presenza di

eteroschedasticità nei residui del modello. Questi problemi sono determinati dal fatto che il modello di regressione lineare non è in grado di catturare correttamente l'autocorrelazione presente nei dati. Si vedrà nella sezione successiva come altri modelli siano, invece, capaci di integrare nel modello questa componente.

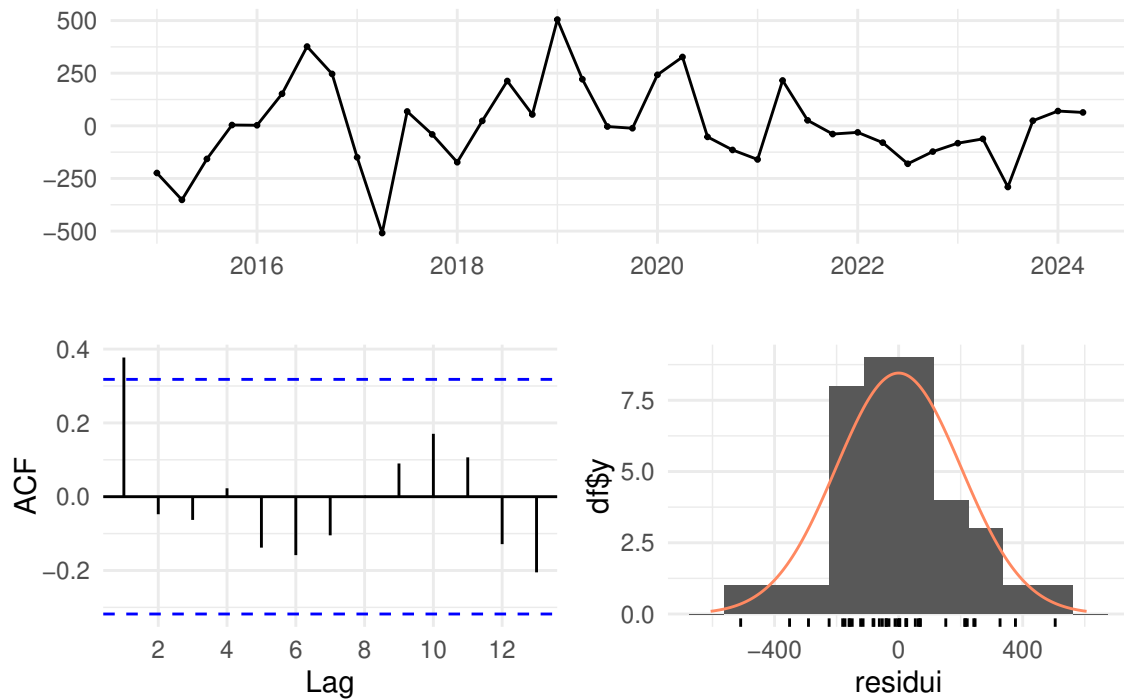


Figura 2.4: in alto la serie dei residui del modello; in basso a sinistra l'ACF e in basso a destra la loro distribuzione.

statistica	valore
errore standard residuo	217.2 su 32 gdl
R^2	0.8736
R^2 corretto	0.8539
statistica F	44.24 su 5 e 32 gdl
p-value	$< 1.885 \times 10^{-13}$
DW test	1.2097

Tabella 2.2: statistiche del modello 2.1.

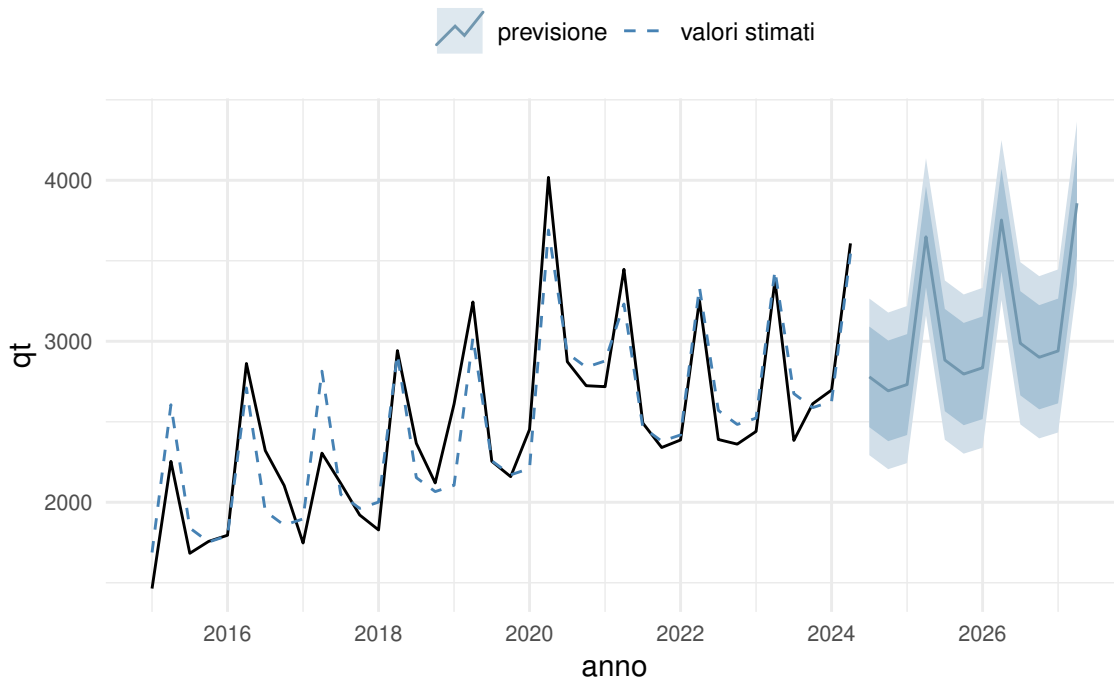


Figura 2.5: grafico delle previsioni di vendita trimestrali di pennelli per i tre anni successivi. La zona scura indica gli intervalli di confidenza di livello 80%, mentre l'area più chiara quelle di livello 95%.

2.1.2 Serie mensile

In questa seconda analisi verrà studiata la serie mensile delle vendite di pennelli. L'obiettivo questa volta è quello di catturare anche la componente di autocorrelazione nei dati, al fine di ottenere previsioni a breve termine molto più precise. Questo è possibile anche grazie ad una disponibilità maggiore di dati osservati.

Analisi esplorativa

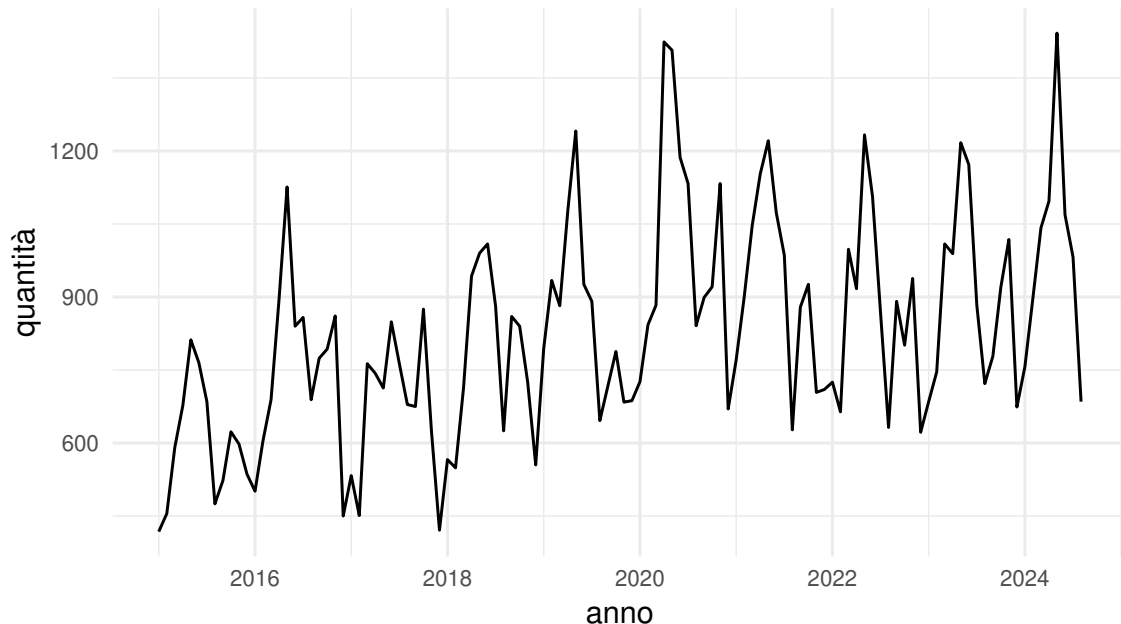


Figura 2.6: grafico delle vendite mensili di pennelli dal 2015.

Dal grafico della serie in figura 2.6 possiamo notare come la serie mensile sia molto più variabile rispetto a quella trimestrale. Questo denota la presenza di alcuni mesi in particolare che registrano costantemente vendite molto elevate o molto scarse durante l'anno. Dal grafico in figura 2.7 è possibile comprendere più efficacemente la distribuzione delle vendite nei vari mesi dell'anno. In ogni boxplot il rettangolo centrale rappresenta il 50% dei dati compresi tra il primo

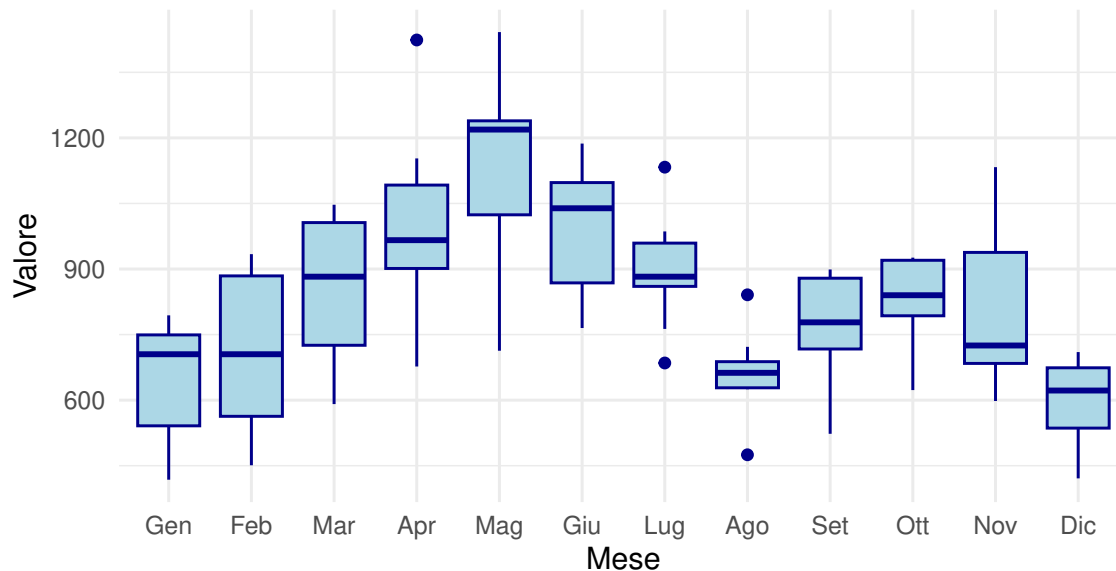


Figura 2.7: boxplot della distribuzione delle vendite di pennelli nei singoli mesi.

quartile e il terzo quartile, mentre la linea blu all'interno del rettangolo rappresenta la mediana. I valori sembrano aumentare progressivamente da gennaio fino a maggio, con un picco evidente in quest'ultimo mese. Nei mesi estivi di giugno e luglio c'è una leggera riduzione, mentre agosto mostra un calo significativo. Negli ultimi mesi dell'anno i valori risalgono leggermente, per poi abbassarsi nuovamente a dicembre. Alcuni mesi, come aprile e maggio, hanno una variabilità molto alta. Al contrario, mesi come luglio e agosto, eccetto qualche outlier, sono caratterizzati dalla presenza di valori molto vicini tra loro.

Nel grafico 2.8 è invece possibile identificare l'andamento delle vendite per i singoli mesi. In questo caso la linea blu rappresenta la media mensile. Da questo grafico è possibile notare ancora più chiaramente la presenza di oscillazioni importanti, in particolare ad aprile, maggio e novembre, frutto anche dell'effetto della pandemia.

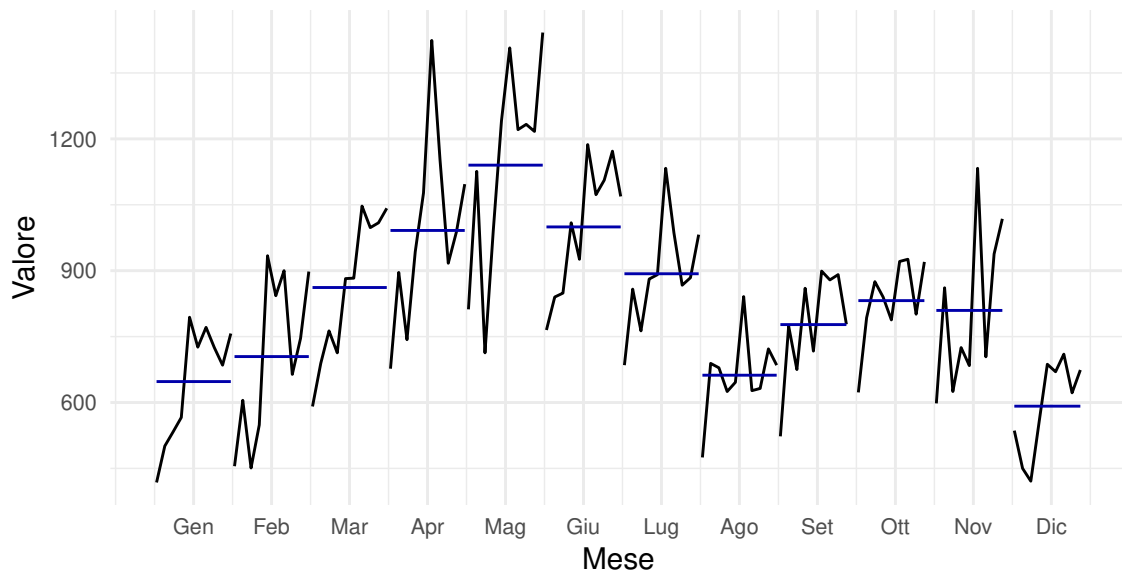


Figura 2.8: grafico stagionale per sotto-serie delle vendite mensili di pennelli.

Modelli di previsione

Per la serie mensile di vendita dei pennelli sono stati adattati tre modelli differenti:

1. Modello di lisciamiento esponenziale di Holt-Winters smorzato con stagionalità moltiplicativa.
2. Modello SARIMA stimato a partire dall'ACF e dalla PACF della serie.
3. Modello SARIMAX stimato con la funzione *auto.arima* di R, includendo una variabile di regressione per descrivere l'influenza della pandemia sulle vendite.

1. Modello di Holt-Winters con trend smorzato

Il modello di Holt-Winters, a differenza degli altri modelli di lisciamiento, consente di catturare la presenza di stagionalità nei dati. La sua funzione di previsione con trend smorzato è la seguente:

$$\begin{aligned}
l_t &= \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(l_{t-1} + \phi b_{t-1}) \\
b_t &= \beta(l_t - l_{t-1}) + (1 - \beta)\phi b_{t-1} \\
S_t &= \gamma \frac{Y_t}{(l_{t-1} + \phi b_{t-1})} + (1 - \gamma)S_{t-s} \\
F_{t+m} &= [l_t + (\phi + \phi^2 + \dots + \phi^m)b_{t-m}]S_{t-s+m}, \tag{2.2}
\end{aligned}$$

dove l_t , b_t e S_t sono rispettivamente le equazioni per il livello, il trend e la stagionalità, mentre α , β e γ sono le costanti di lisciamiento associate alle tre componenti. Il parametro ϕ , invece, è la costante di smorzamento. In questa formulazione la stagionalità è moltiplicativa, ma è bene specificare che questa può essere anche additiva. Si tende ad utilizzare la seconda solo nel caso in cui l'ampiezza delle variazioni stagionali sia costante nel tempo.

parametro	stima
alpha	0.3283
beta	0.0004
gamma	0.0004
phi	0.9735
componente	valore di inizializzazione
livello	516.7811
trend	13.6149
stagionalità	0.706; 0.9839; 1.0114; 0.9514; 0.7996; 1.0878; 1.2042; 1.3728; 1.189; 1.0513; 0.8518; 0.7909;

Tabella 2.3: stima dei parametri e dei valori di inizializzazione del modello 2.2.

A partire dalle stime riportate nella tabella 2.3 è possibile fare diverse considerazioni. Le costanti di lisciamiento di trend e stagionalità, a

differenza di quella per il livello, presentano dei valori molto piccoli; questo significa che i valori di trend e stagionalità tendono a rimanere stabili nel corso della serie. Tuttavia, il valore iniziale del trend è piuttosto elevato ed indica una crescita iniziale significativa. Infine, il parametro di smorzamento ha un valore molto vicino ad 1, il che significa che il trend ha un'evoluzione molto graduale.

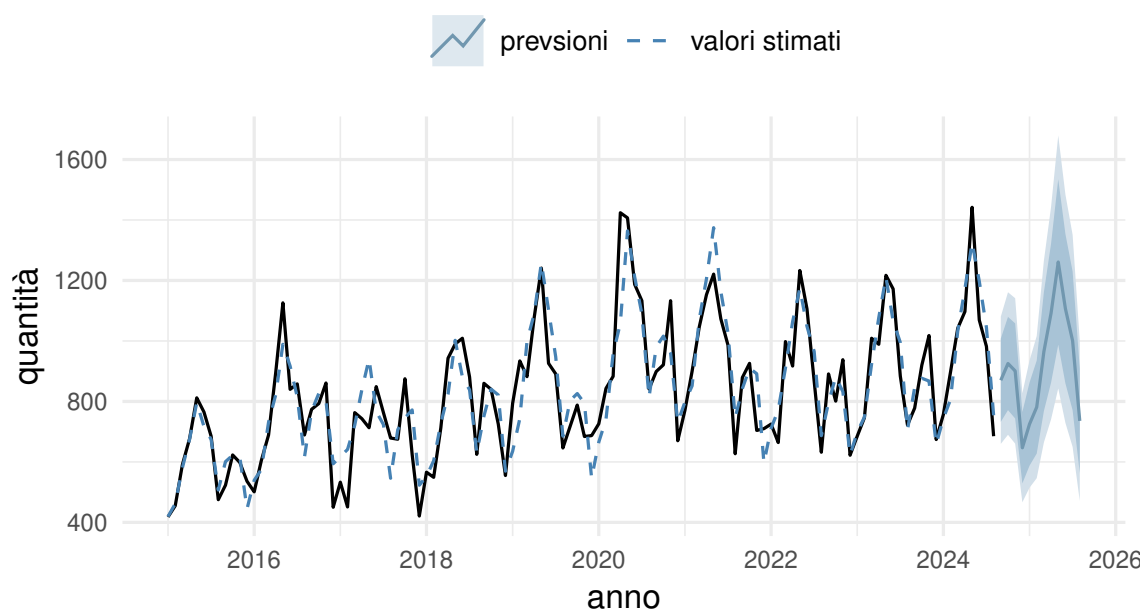


Figura 2.9: grafico delle previsioni del modello di Holt-Winters smorzato per le vendite mensili dei pennelli. La zona più scura indica gli intervalli di previsione all'80%; le zone più chiare rappresentano gli intervalli di confidenza al 95%.

2. Modello SARIMA

Il modello SARIMA (Seasonal Autoregressive Integrated Moving Average) è un'estensione del modello ARIMA che include anche componenti stagionali. Analizzando i grafici delle funzioni di autocorrelazione del modello in figura 2.10 e successivamente il grafico

della serie differenziata con ritardo stagionale in figura 2.11, è stato scelto un modello SARIMA(3,0,0)(0,1,1)₁₂ con drift. Di seguito l'equazione del modello nella notazione backshift:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B^{12})y_t = c + (1 - \Theta_1 B^4)\epsilon_t \quad (2.3)$$

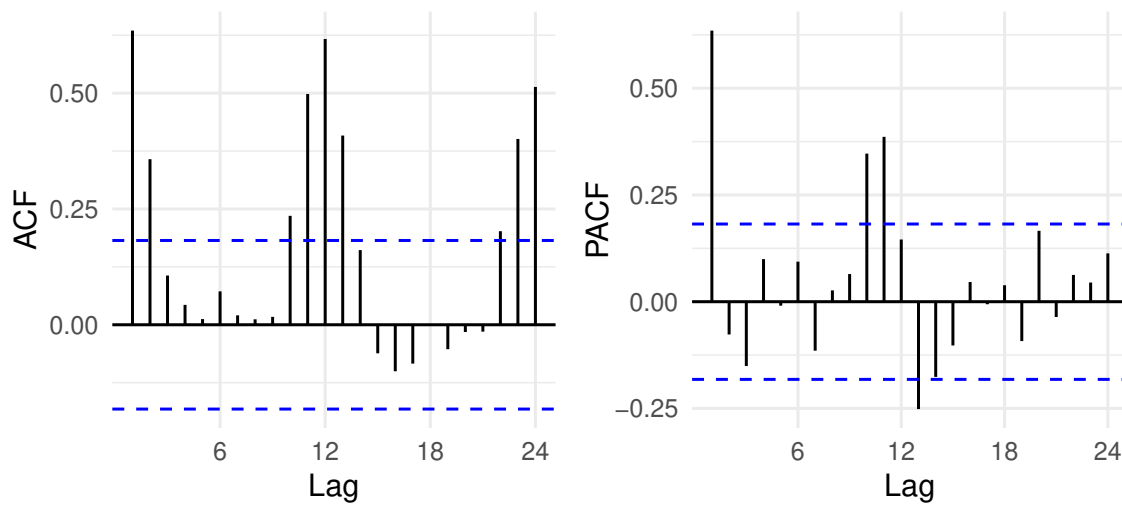


Figura 2.10: funzione di autocorrelazione e di autocorrelazione parziale della serie mensile originale.

	AR1	AR2	AR3	SMA1	drift
coefficiente	0.2112	0.2386	0.2518	-0.7930	2.9707
errore standard	0.0956	0.0943	0.0961	0.1216	0.9853

Tabella 2.4: stima dei parametri del modello 2.3.

Dalle stime dei coefficienti del modello riportate nella tabella 2.4, è possibile notare come i coefficienti autoregressivi (AR1, AR2, AR3) siano tutti positivi. La serie sembra quindi mostrare una dipendenza temporale persistente nei dati, con una memoria che si estende fino a

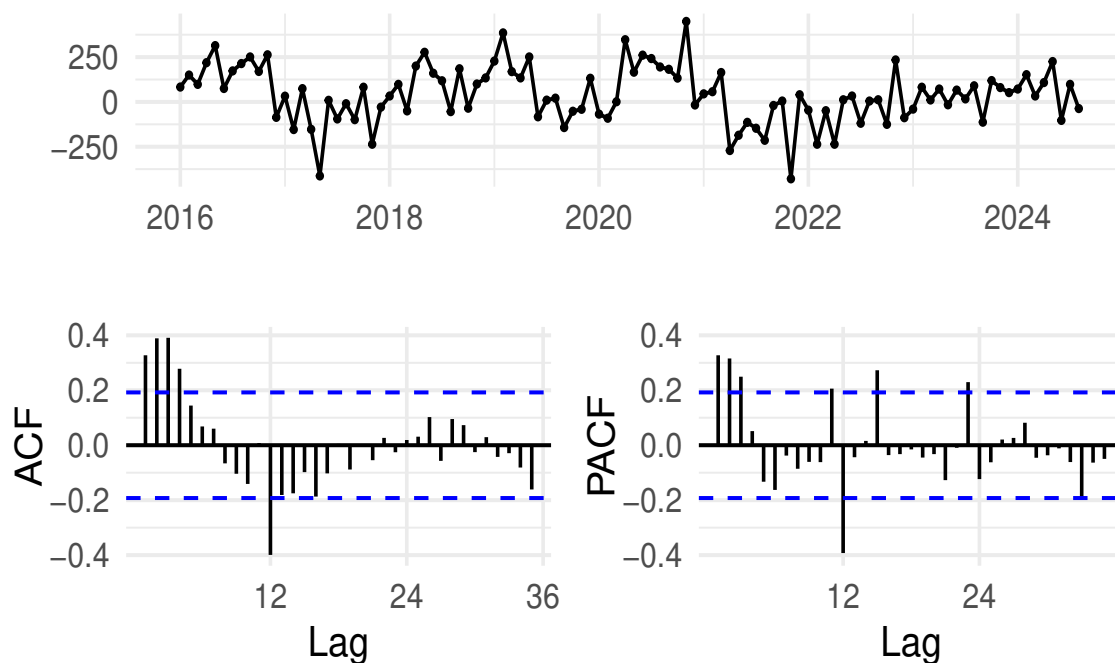


Figura 2.11: in alto la serie differenziata con ritardo stagionale; in basso l'ACF e la PACF.

tre periodi passati, di intensità progressivamente maggiore. La componente a media mobile stagionale (SMA1) è significativa e negativa, indicando che gli errori del periodo precedente influenzano negativamente i valori futuri della serie. Infine, il termine di drift denota un trend positivo, ma il valore elevato per il suo errore standard evidenzia come questa stima potrebbe non essere del tutto affidabile. A questo punto, per poter valutare il modello, è necessario condurre un'analisi diagnostica dei residui. Dall'analisi grafica in figura 2.12 sembra che i residui si comportino come un white noise; questa ipotesi è confermata dal test di Ljung-Box, i cui risultati sono riportati nella tabella 2.5.

Test di Ljung-Box	
statistica	18.220
gdl	23
p-value	0.745

Tabella 2.5: risultati del test di Ljung-Box per testare l'incorrelazione dei residui del modello SARIMA.

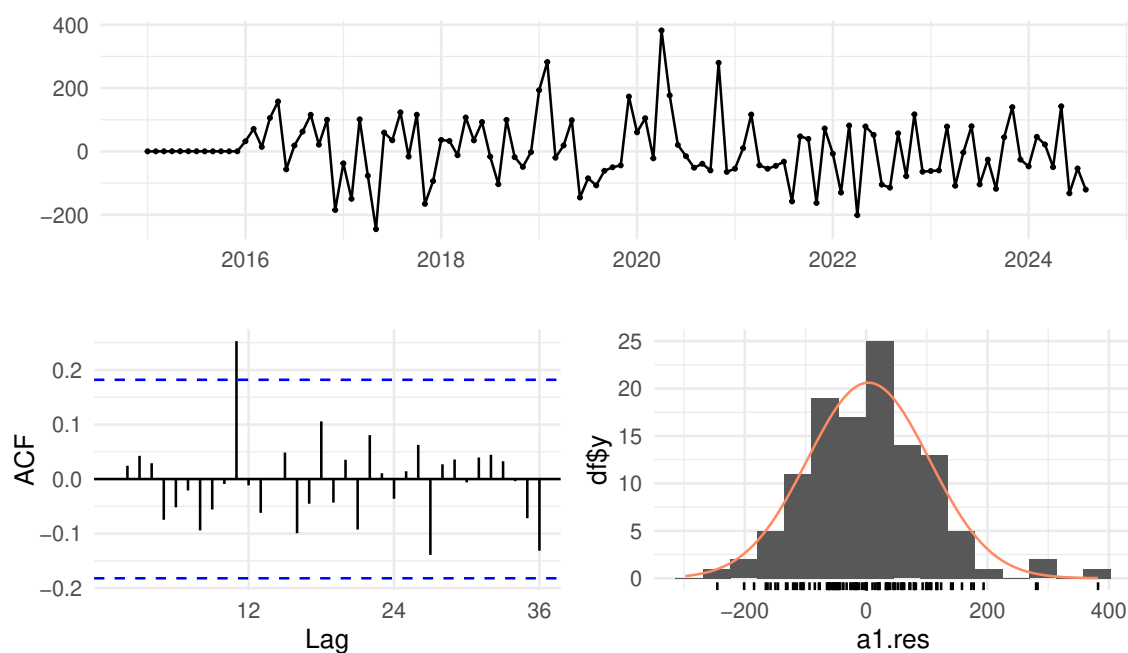


Figura 2.12: in alto la serie dei residui del modello; in basso a sinistra l'ACF e in basso a destra la loro distribuzione.

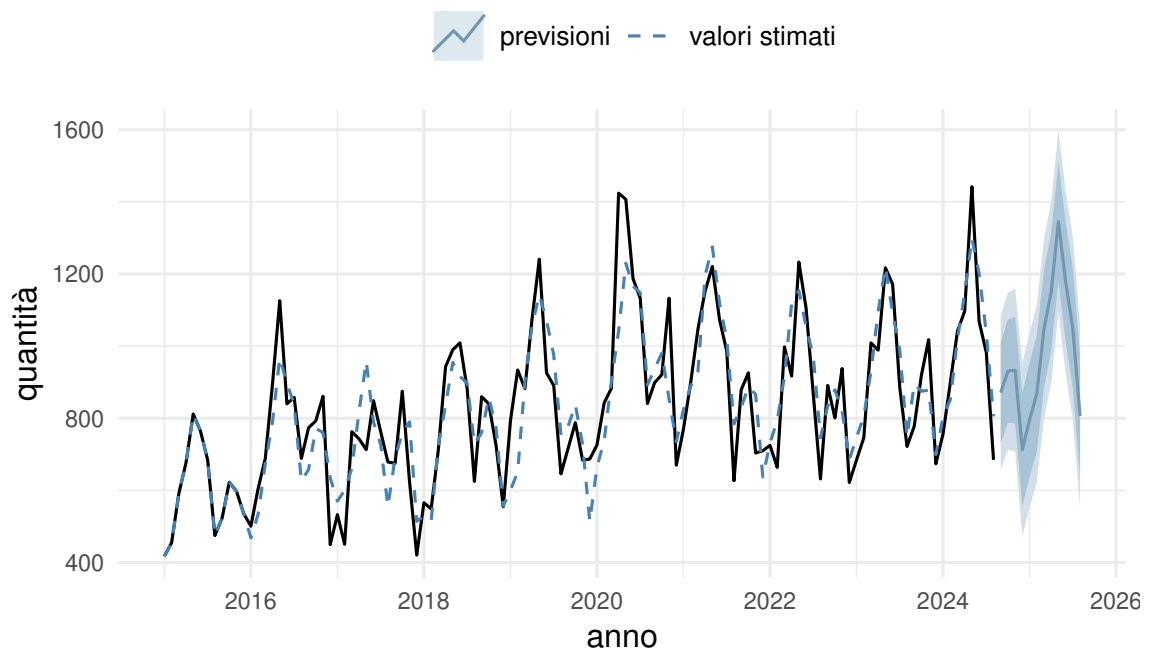


Figura 2.13: grafico delle previsioni del modello SARIMA(3,0,0)(0,1,1)₁₂ per le vendite mensili di pennelli. La zona più scura indica gli intervalli di previsione all'80%; le zone più chiare rappresentano gli intervalli di confidenza al 95%.

3. Modello SARIMAX

In quest'ultimo modello, stimato dalla funzione *auto.arima* della libreria *forecast* di R, è stata inserita, come nel caso del modello di regressione per la serie trimestrale, una variabile dummy per descrivere l'influenza della pandemia sulle vendite. Il modello che minimizza l'AIC è un SARIMAX(0,0,3)(0,1,1)₁₂ con drift:

$$(1 - B^4)y_t = c + \lambda x_t + (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)(1 - \Theta_1 B^4)\epsilon_t, \quad (2.4)$$

	MA1	MA2	MA3	SMA1	drift	xreg
coefficiente	0.1397	0.2074	0.2636	-0.7836	2.7998	184.8456
errore standard	0.0989	0.0939	0.0894	0.1214	0.4887	39.8794

Tabella 2.6: stima dei parametri del modello 2.4.

Alla luce delle considerazioni fatte per il modello stimato in precedenza, è possibile fare un confronto. Nel modello precedente i coefficienti della parte autoregressiva erano positivi e leggermente più grandi rispetto a quelli dei termini MA. Tuttavia, entrambi i modelli suggeriscono una dipendenza dai valori passati, con un'influenza crescente dei termini più lontani. I coefficienti SMA1 nei due modelli sono praticamente uguali, rispettivamente -0.7836 e -0.7930. Entrambi i modelli, quindi, mostrano una forte correlazione negativa tra gli errori passati stagionali e i valori correnti, confermando la presenza significativa di stagionalità. Il termine di drift, invece, pur essendo leggermente inferiore rispetto a quello del modello SARIMA, ha errore standard decisamente più contenuto. Infine, il termine di regressione (184.845) suggerisce come gli effetti della pandemia abbiano avuto un'influenza significativa e positiva sui valori della serie temporale.

Anche in questo caso si accetta l'ipotesi nulla di incorrelazione seriale dei residui del test di Ljung-Box (tabella 2.7) con un livello di significatività molto elevato.

Test di Ljung-Box	
statistica	16.724
gdl	23
p-value	0.822

Tabella 2.7: Risultati del test di Ljung-Box per testare l'incorrelazione dei residui nel modello SARIMAX.

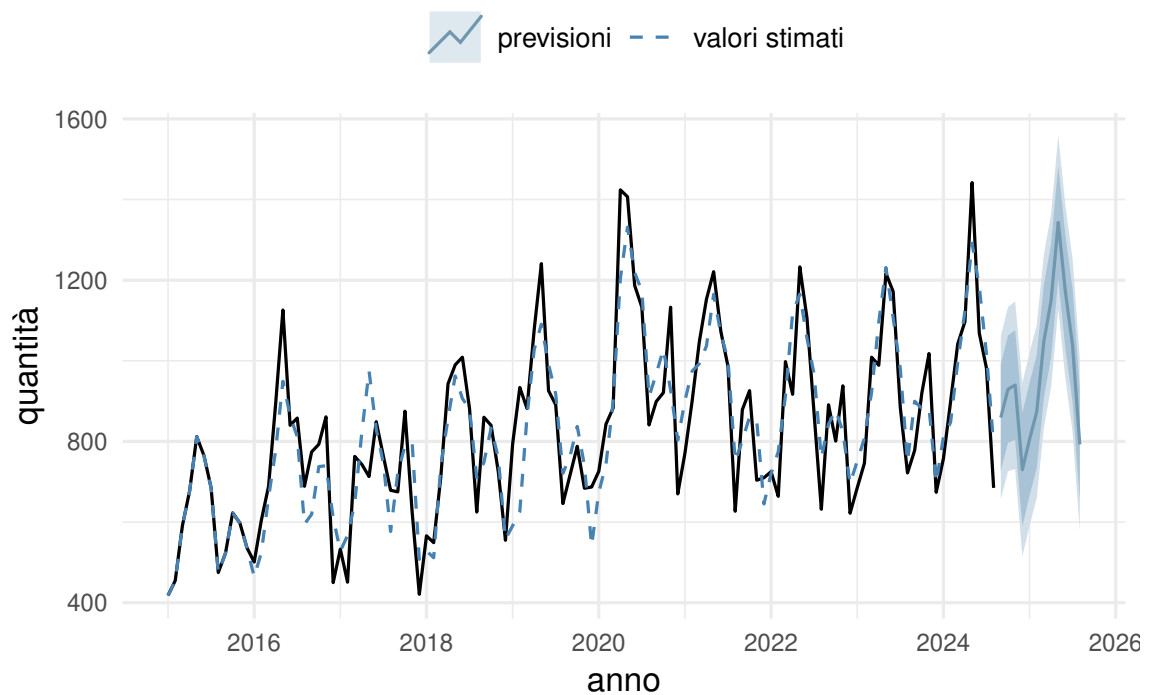


Figura 2.14: grafico delle previsioni del modello $ARIMAX(0,0,3)(0,1,1)_{12}$ per le vendite mensili di pennelli. La zona più scura indica gli intervalli di previsione all'80%; le zone più chiare rappresentano gli intervalli di confidenza al 95%.

Scelta del modello

A questo punto dell'analisi occorre effettuare una valutazione per poter scegliere uno dei tre modelli proposti. Un primo confronto può essere svolto attraverso i criteri di informazione AIC e BIC. Dai valori in tabella 2.8 è chiaro come gli ultimi due modelli abbiano un adattamento notevolmente superiore al primo. Il modello SARIMAX ha l'adattamento migliore.

	AIC	BIC
Holt-Winters	1636.044	1685.609
SARIMA(3,0,0)(0,1,1) ₁₂	1288.97	1304.84
SARIMAX(0,0,3)(0,1,1) ₁₂	1277.95	1296.46

Tabella 2.8: Criteri di informazione per i tre modelli

Per verificare la validità previsiva dei tre modelli è stata poi condotta un'analisi di accuratezza delle previsioni tramite campione di stima e campione test. In questo caso i modelli sono stati stimati utilizzando i dati fino ad agosto 2023.

Campione di stima					
	ME	RMSE	MAE	MPE	MAPE
Holt-Winters	-2.784	93.545	73.829	-1.537	9.382
SARIMA(3,0,0)(0,1,1) ₁₂	4.137	96.575	71.720	-0.725	8.954
SARIMAX(0,0,3)(0,1,1) ₁₂	6.518	91.766	70.842	-0.316	8.892

Tabella 2.9: statistiche di accuratezza dei tre modelli sul campione di stima

Come si evince dai valori nelle tabelle 2.9 e 2.10 il modello SARIMAX è generalmente il modello con le migliori misure di accuratezza delle previsioni.

Campione di test					
	ME	RMSE	MAE	MPE	MAPE
Holt-Winters	51.333	98.789	72.246	4.626	7.280
SARIMA(3,0,0)(0,1,1) ₁₂	-12.550	91.234	77.143	-2.595	8.284
SARIMAX(0,0,3)(0,1,1) ₁₂	-2.097	87.416	73.066	-1.510	7.768

Tabella 2.10: statistiche di accuratezza dei tre modelli sul campione di test

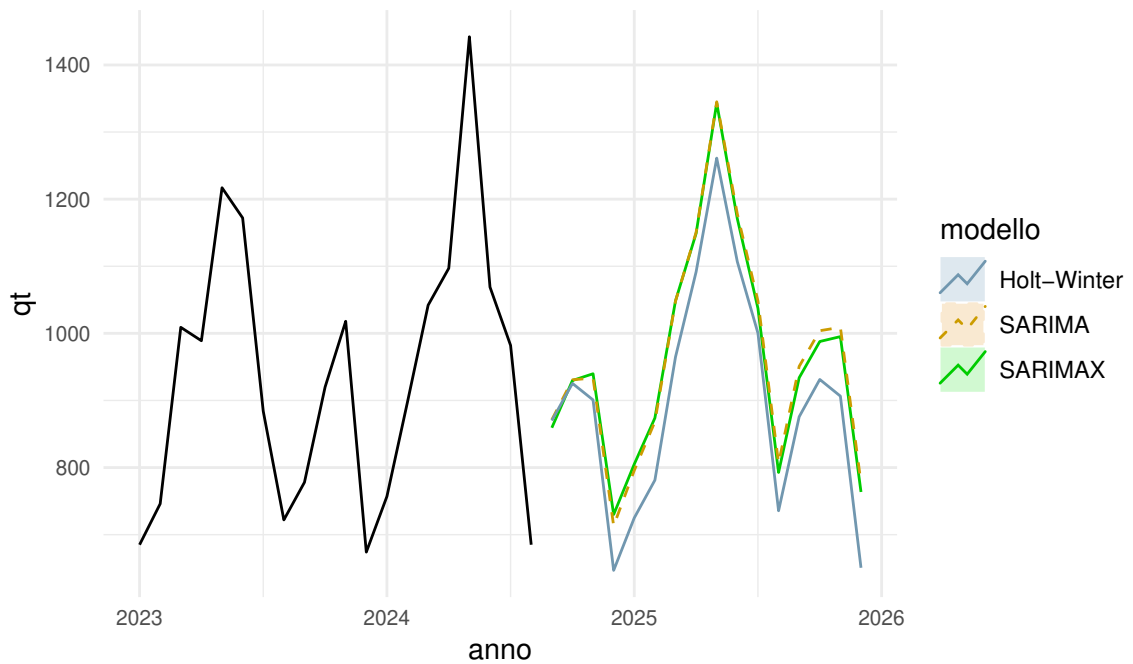


Figura 2.15: zoom sul confronto delle previsioni a breve termine per i tre modelli.

Capitolo 3

Gestione del rischio di obsolescenza

Con il rapido evolversi della tecnologia e delle preferenze dei consumatori, alcuni prodotti possono perdere rapidamente valore, trasformando l'eccesso di scorte in una perdita economica. Per questo motivo, le aziende devono adottare strategie efficaci per ridurre il rischio di obsolescenza, integrando modelli di gestione del ciclo di vita del prodotto nelle loro analisi statistiche.

3.1 Caso studio: duplicati di chiavi

Nel seguente caso studio, verrà analizzata la serie storica delle vendite di chiavi a profilo universale duplicate. Questo tipo di chiave è stato il più utilizzato negli ultimi decenni; tuttavia, per quanto riguarda la sicurezza, risulta ormai obsoleto. Altre tipologie di chiavi, come le chiavi a doppia mappa e le chiavi punzonate sono, infatti, molto più valide. Inoltre, l'evoluzione tecnologica dei sistemi di sicurezza e serratura porterà nei prossimi anni un graduale calo nell'utilizzo delle chiavi in generale.

Analisi esplorativa

Dal grafico in figura 3.1 si possono fare un paio di considerazioni:

- le vendite sono caratterizzate da un notevole incremento tra l'anno 2015 e l'anno 2020, per poi tornare stabili negli anni successivi.
- la componente stagionale è piuttosto evidente per tutta la serie, anche se l'altezza dei picchi periodici non è molto costante.

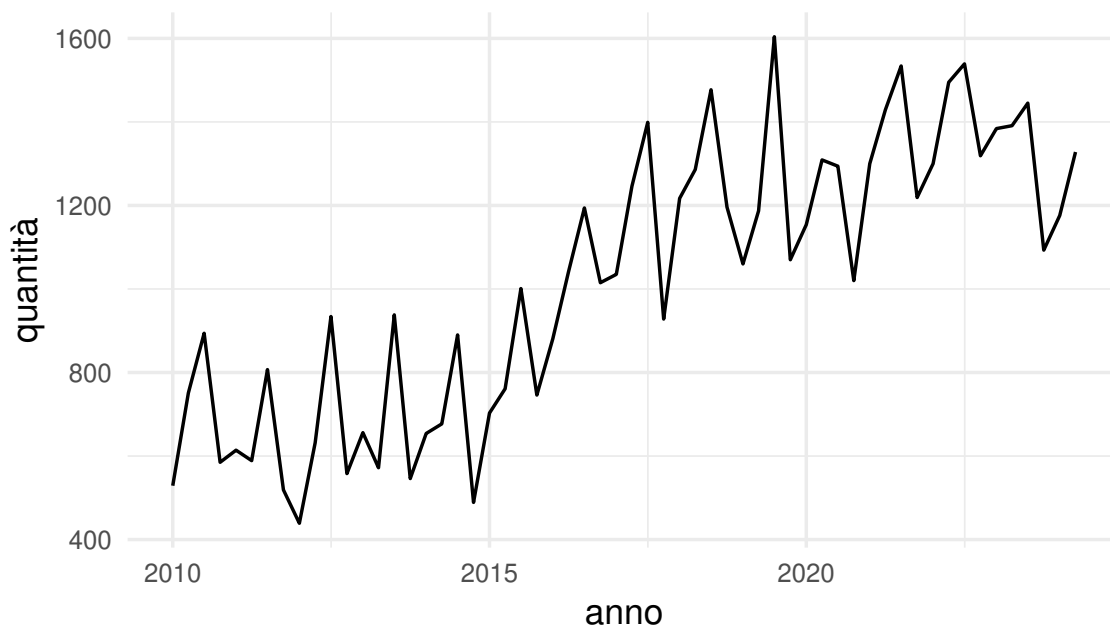


Figura 3.1: grafico della serie storica trimestrale delle vendite di chiavi a profilo universale duplicate.

Il grafico in figura 3.2 aiuta a comprendere meglio la stagionalità della serie. Il terzo trimestre presenta sicuramente valori in media superiori agli altri periodi. Questa circostanza è determinata dalla posizione in cui si trova l'attività commerciale, ovvero una zona con molte seconde case abitate solamente nel periodo estivo.

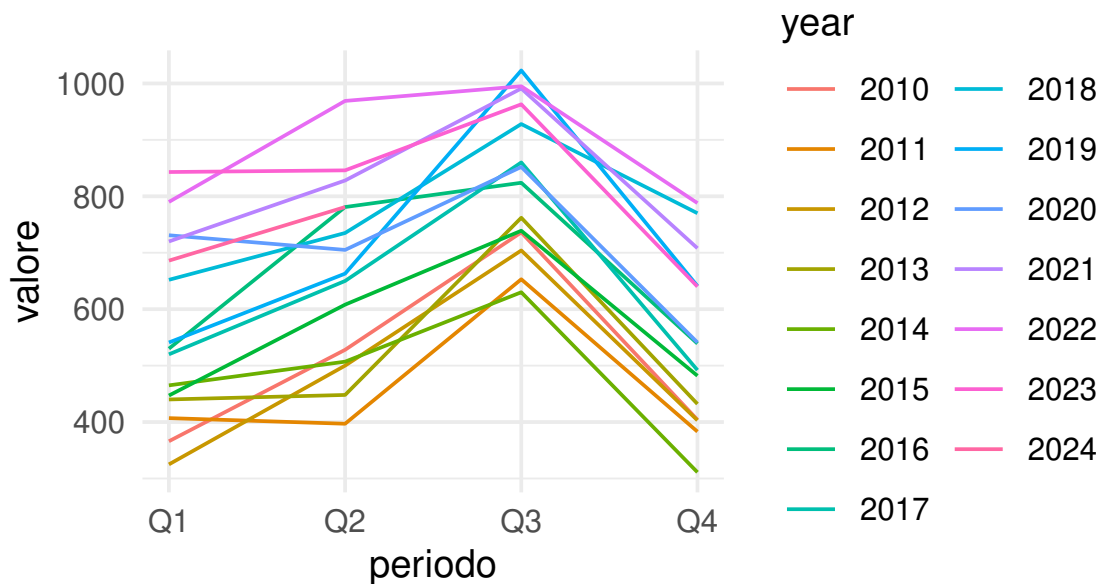


Figura 3.2: grafico dei valori per trimestre delle vendite di chiavi a profilo universale.

Modelli di previsione

Per questo caso studio verranno proposti due differenti modelli di diffusione:

- Modello di Bass Generalizzato con shock esponenziale.
- Modello di Bass Standard con errori SARIMA (Modello SARIMAX).

1. Modello di Bass generalizzato con shock esponenziale

Per poter effettuare una previsione a lungo termine e determinare l'andamento delle vendite nelle fasi finali del ciclo di vita delle chiavi a profilo universale è stato stimato un modello di Bass generalizzato con uno shock esponenziale (Guidolin 2023). Il modello è il seguente:

$$z(t) = m \frac{1 - e^{-(p+q) \int_0^t x(\tau) d\tau}}{1 + \frac{q}{p} e^{-(p+q) \int_0^t x(\tau) d\tau}}, \quad (3.1)$$

con

$$x(t) = 1 + c_1 e^{b_1(t-a_1)} I_{t \geq a_1} \quad (3.2)$$

	stima	errore st.	p-value
<i>m</i>	122328.22	34879.50622	0.00094
<i>p</i>	0.00517	0.00148	0.00098
<i>q</i>	0.01181	0.00230	0.00000
<i>a1</i>	24.10948	0.36983	0.00000
<i>b1</i>	0.01864	0.01091	0.09360
<i>c1</i>	0.52683	0.04218	0.00000

Tabella 3.1: stime dei parametri del modello di Bass generalizzato con shock esponenziale.

La stima del mercato potenziale (*m*) è di circa 122000 unità, un valore piuttosto elevato; questo è spiegato dal fatto che il prodotto in questione si trova già in una fase avanzata del suo ciclo di vita. Questo aspetto è confermato anche dal valore molto piccolo del parametro (*p*), che rappresenta la propensione degli innovatori a iniziare l'adozione del prodotto senza influenze esterne. Il parametro (*q*), infine, indica una crescita moderata per effetto dell'adozione del prodotto da parte degli imitatori. In tutti e tre i casi i parametri sono altamente significativi.

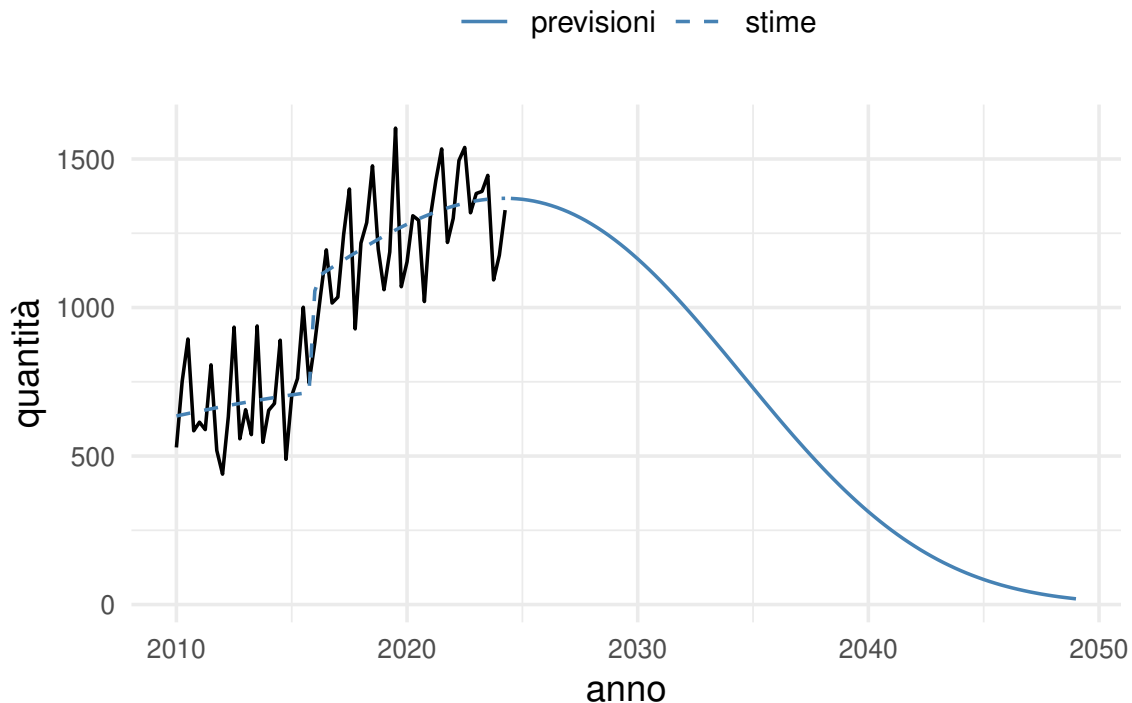


Figura 3.3: grafico delle previsioni a lungo termine per la vendita di chiavi a profilo universale del modello 3.1.

2. Modello di Bass Standard con errori SARIMA

Il modello di Bass è costruito per catturare il comportamento medio non lineare di una serie, tralasciando altre componenti come autocorrelazione e stagionalità. Dunque, nel momento in cui ci si trova di fronte a serie storiche stagionali e autocorrelate, può essere utile integrare il modello di Bass per ottenere previsioni a breve termine più precise. Un possibile approccio per risolvere questa problematica è l'utilizzo come regressore all'interno di un modello Arimax dei valori stimati dal modello di Bass. Il processo di specificazione e stima di un modello di questo tipo è il seguente (Guidolin 2023):

1. selezionare un modello di diffusione non lineare come il modello di Bass o Bass generalizzato.

2. condurre un'analisi diagnostica dei residui.
3. se è presente autocorrelazione nei residui, includere le stime del modello non lineare come regressore in un modello ARIMAX.
4. fare una selezione del modello tramite AIC, controllando sempre le stime dei parametri, in particolare la stima della regressione.

In questo caso è stato scelto attraverso il criterio di Akaike il seguente modello SARIMAX(3,0,0)(0,1,1)₄:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B^4)y_t = \lambda x_t + (1 - \Theta_1 B^4)\epsilon_t, \quad (3.3)$$

con ϵ_t processo white noise e x_t covariata al tempo t , ottenuta a partire dal modello di Bass così definito:

$$y(t) = m \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}} \quad (3.4)$$

	AR1	AR2	AR3	SMA1	λ
coefficiente	0.4360	0.1109	0.2187	-0.9742	0.8723
errore standard	0.1332	0.1534	0.1381	0.6701	0.1484

Tabella 3.2: stima dei parametri del modello 3.3.

Le stime dei parametri del modello sono riportate nella tabella 3.2. Il modello ha una forte componente autoregressiva, specialmente al primo e al terzo ritardo, e una significativa componente a media mobile stagionale. Il coefficiente della regressione è particolarmente elevato e ha un errore standard contenuto.

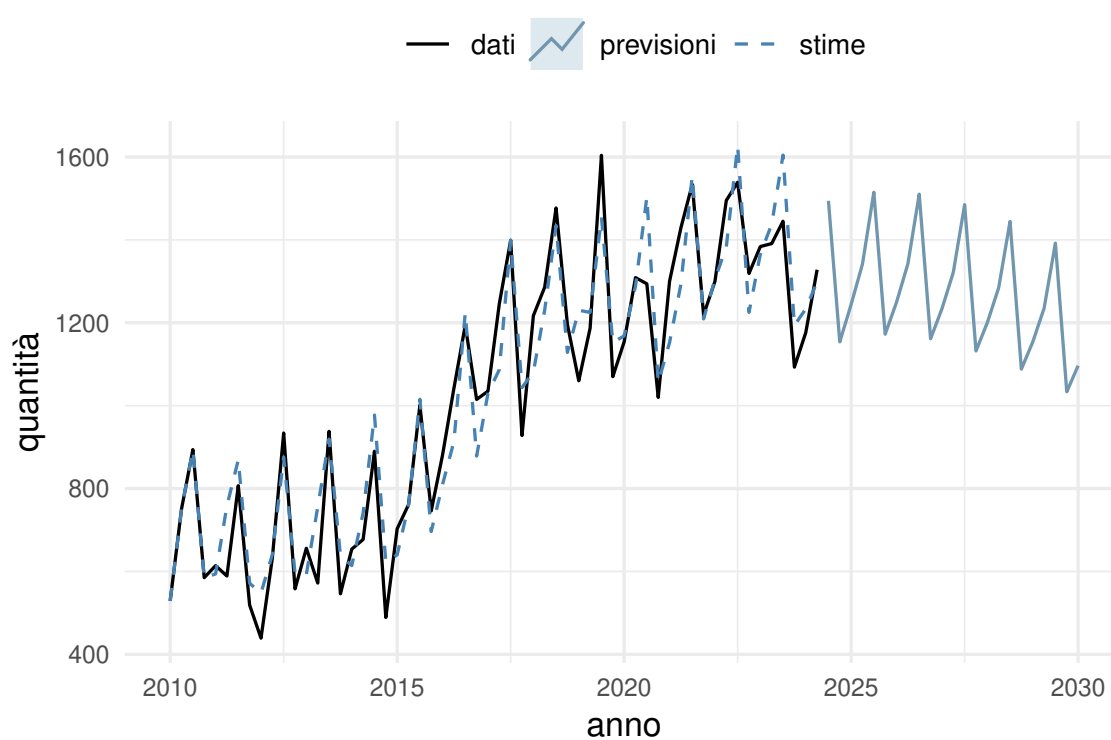


Figura 3.4: grafico delle previsioni a medio termine del modello 3.3 per le vendite di duplicati di chiavi a profilo universale.

Capitolo 4

Analisi delle tendenze di mercato

Analizzare e prevedere le tendenze nel mercato di riferimento è un aspetto cruciale per le attività commerciali (Paci 2019). Di seguito sono elencati i principali fattori che influenzano il settore del Fai Da Te:

1. Cicli economici; durante i periodi di crescita economica, le vendite di vernici e attrezzature per il Fai Da Te tendono ad aumentare, poiché i consumatori e le imprese investono in progetti di ristrutturazione e manutenzione. Al contrario, durante le recessioni, i consumatori potrebbero ridurre la spesa per progetti non essenziali, influenzando negativamente le vendite. Inoltre, questo settore dipende particolarmente da altri segmenti di mercato come quello edile o immobiliare.
2. Comportamento e preferenze dei consumatori; negli ultimi anni, c'è stato un aumento dell'interesse per il Fai Da Te, alimentato anche dalla diffusione di piattaforme online che offrono tutorial e ispirazioni. Allo stesso tempo i consumatori sono sempre più orientati verso prodotti ecologici e con impatto ambientale ridotto.
3. Normative ambientali; le regolamentazioni, specialmente quelle riguardanti i VOC (composti organici volatili) e l'impatto

ambientale delle vernici influenzano notevolmente la produzione e la vendita. Le attività commerciali devono essere aggiornate sulle normative locali e nazionali per garantire la conformità dei prodotti.

4. Innovazioni tecnologiche: lo sviluppo e la vendita di nuovi prodotti più efficaci e innovativi consente alle aziende di mantenere la competitività sul mercato. Questo processo richiede investimenti in ricerca e formazione del personale. Un aspetto importante legato alle innovazioni tecnologiche è la digitalizzazione delle vendite; le attività commerciali devono implementare strategie di vendita online per raggiungere un pubblico più vasto e migliorare l'esperienza per il cliente.

4.1 Caso studio: smalti per legno e metallo

Come anticipato, le normative ambientali riguardanti i VOC hanno avuto e continueranno ad avere un grosso impatto sulla produzione e sulla vendita delle vernici. I VOC sono composti chimici organici che evaporano facilmente, anche a temperatura ambiente. Questi possono causare danni sia all'ambiente che alla salute delle persone, soprattutto se inalati in quantità elevate. Le normative che regolano produzione, commercializzazione ed utilizzo di questi composti chimici sono la Direttiva europea 199/13/CE, e la Direttiva europea 2004/42/CE (Unione Europea 2004), entrambe intese a tutelare l'ambiente e la salute dei lavoratori. Nel caso studio che verrà affrontato in questa sezione si vedrà come l'introduzione di queste normative abbia interessato e in parte stravolto il mercato degli smalti professionali per legno e metallo. L'analisi di previsione sarà strutturata su due scenari scelti in base all'orizzonte temporale di previsione.

Analisi esplorativa

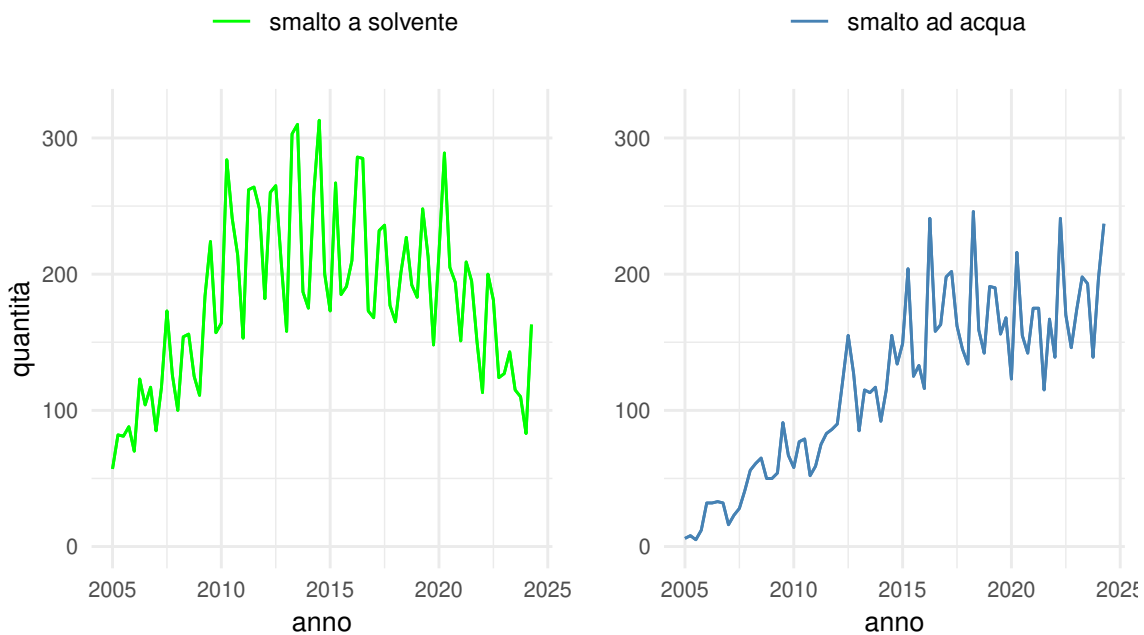


Figura 4.1: serie trimestrale delle vendite di smalti per legno e metallo suddivisa per elemento chimico di composizione del prodotto dal 2005.

Il grafico in figura 4.1 mostra l'andamento delle vendite trimestrali delle due categorie di smalto dal 2005 al 2024. Di seguito sono elencate alcune considerazioni importanti:

- smalto a solvente: la serie presenta un chiaro andamento non lineare; il grafico evidenzia, infatti, un trend crescente fino al 2015 circa, con un picco che si mantiene per alcuni anni, seguito poi da un calo significativo. Il trend decrescente a partire dal 2016 riflette l'effetto delle normative ambientali più stringenti sui VOC, che, sebbene siano state introdotte già dal 2004, hanno avuto conseguenze negative importanti sulle vendite di prodotti a solvente solo dopo diversi anni. Inoltre, per tutta la serie, è evidente la presenza di picchi stagionali più o meno marcati.

- smalto all'acqua: in questo caso il grafico mostra un trend di crescita più o meno costante lungo l'intero periodo. Lo shock positivo iniziale di vendite coincide quasi perfettamente con le prime restrizioni sui prodotti a solvente. Prima di queste normative, infatti, le vernici a base acqua avevano una percentuale di mercato molto ridotta. Come nell'altra serie, anche le vendite di vernici all'acqua sono influenzate da una componente stagionale, più evidente dal 2015 in avanti.

In sintesi, è chiara la presenza di due tendenze contrapposte: il declino degli smalti a solvente e la crescita di quelli all'acqua.

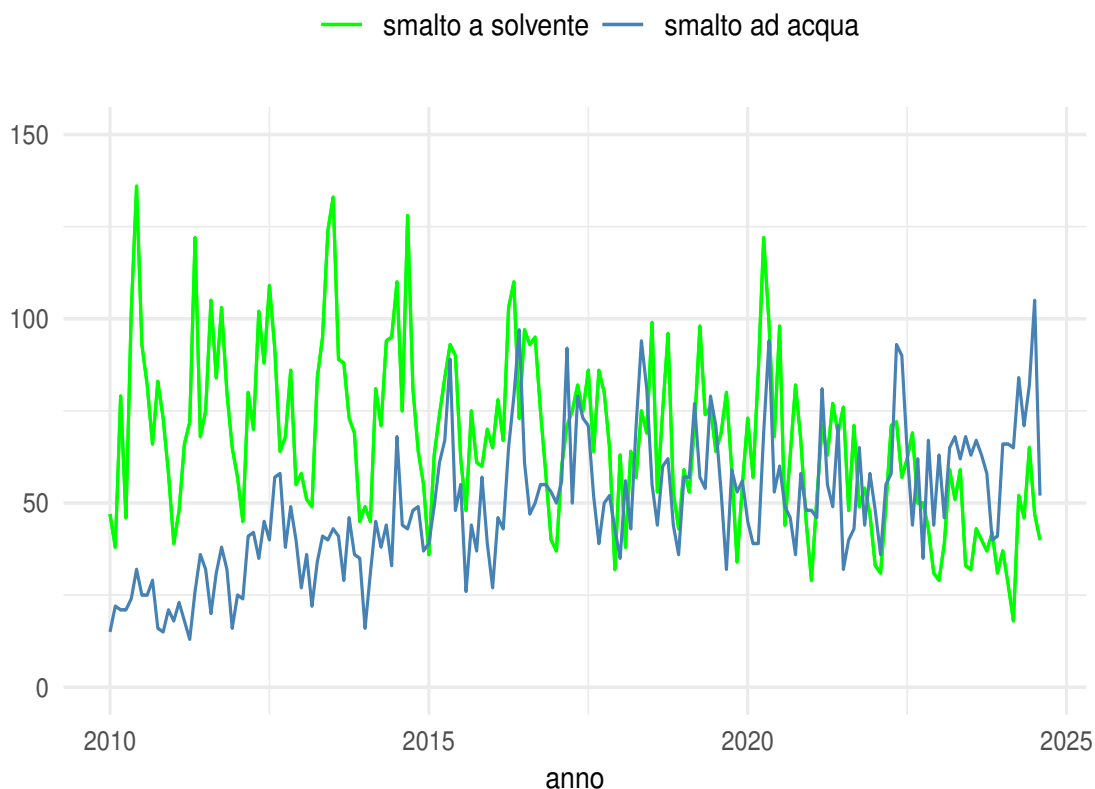


Figura 4.2: serie mensile delle vendite di smalti per legno e metallo suddivisa per elemento di composizione del prodotto dal 2010.

Il grafico in figura 4.2 aiuta a comprendere meglio l'inversione di tendenza tra i due prodotti negli ultimi anni.

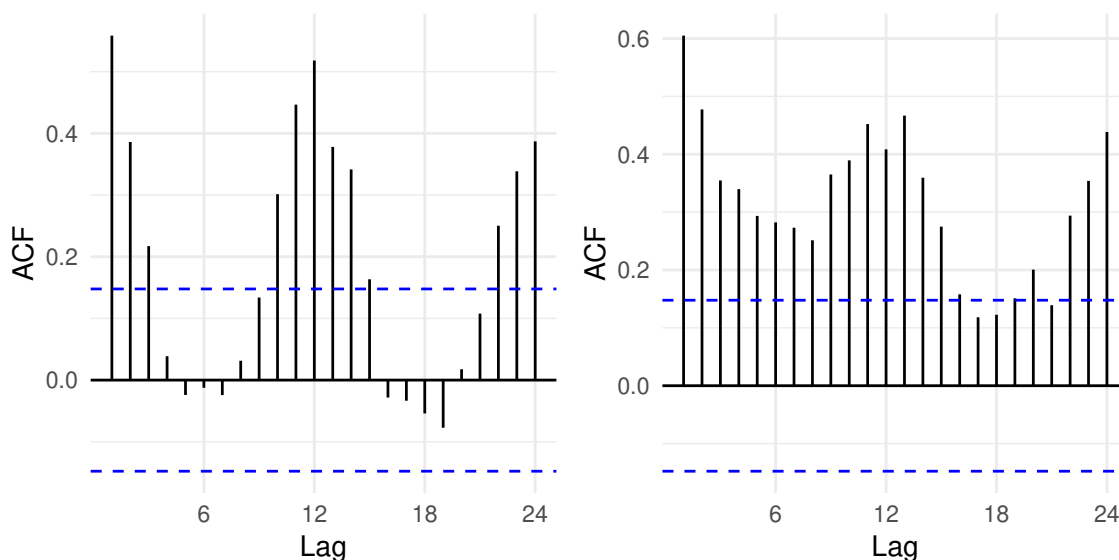


Figura 4.3: funzione di autocorrelazione; a sinistra per le vendite di vernici a solvente, a destra per le vendite di vernici ad acqua

I due correlogrammi in figura 4.3 confermano la presenza di trend in entrambe le serie, ma in modo particolare nella serie delle vendite di vernici a base acqua. Inoltre, le due serie presentano picchi periodici significativi.

4.1.1 Scenari di previsione

Nel contesto di un'analisi di previsione verranno proposti due scenari differenti:

1. **Scenario di previsione a lungo termine:** questo scenario assume che i limiti per il contenuto di VOC nelle vernici vengano ulteriormente abbassati, provocando un declino abbastanza rapido delle vendite di vernici a solvente. I modelli che verranno stimati sono:

- (a) Modello di Bass Standard per le vernici a solvente.
- (b) Modello di Holt con trend smorzato per le vernici ad acqua

2. **Scenario di previsione a medio-breve termine:** in questo caso non vengono fatte ipotesi particolari riguardanti le normative sui VOC. Dunque si assume che, almeno nel breve periodo, non vengano poste ulteriori restrizioni. Per questo scenario verranno stimati due diversi modelli della classe SARIMA.

1. Scenario di previsione a lungo termine

Dal momento che l'attenzione verso la sostenibilità ambientale e la sicurezza sul lavoro è aumentata notevolmente negli ultimi anni, sarebbe lecito aspettarsi che a breve verranno promulgate nuove restrizioni legate al contenuto di VOC nei prodotti vernicianti. In questo caso si ridurrebbe di molto l'utilizzo di vernici e smalti a solvente, mentre i prodotti ad acqua continuerebbero a guadagnare spazio nel mercato. A partire da queste considerazioni sono stati scelti due modelli, uno per ciascuna serie di vendite. Nel caso degli smalti a solvente è stato scelto un modello di Bass Standard, affinché venisse descritto in modo adeguato il declino rapido delle vendite. Il modello stimato è il seguente:

$$z(t) = m \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}} \quad (4.1)$$

	stima	errore st.	p-value
m	16489.27	157.4145	4.53×10^{-83}
p	0.0052	6.94×10^{-05}	1.13×10^{-72}
q	0.0473	0.0010	1.40×10^{-57}

Tabella 4.1: stime dei parametri del modello di Bass Standard 4.1.

Per la serie delle vernici ad acqua, invece, è stato stimato il seguente modello di lisciamiento esponenziale di Holt con trend smorzato:

$$\begin{aligned}
 l_t &= \alpha Y_t + (1 - \alpha)(l_{t-1} + \phi b_{t-1}) \\
 b_t &= \beta(l_t - l_{t-1}) + (1 - \beta)\phi b_{t-1} \\
 F_{t+m} &= l_t + (\phi + \phi^2 + \dots + \phi^m)b_t.
 \end{aligned}
 \tag{4.2}$$

parametro	stima
α	0.0612
β	0.0592
ϕ	0.9263
componente valore di inizializzazione	
livello	2.67
trend	5.2201

Tabella 4.2: stima dei parametri e dei valori di inizializzazione del modello 4.2.

Dal grafico in figura 4.4 si può vedere come le previsioni per le vernici ad acqua descrivano una crescita iniziale molto significativa, che si fermerà fino a coprire una buona parte del mercato potenziale. Nel lungo termine, quindi, la crescita dei prodotti a base acqua tenderà a stabilizzarsi; infatti, è molto probabile che durante questa transizione vengano formulati nuovi solventi di origine biologica che, per certi casi, potrebbero sostituire a loro volta le vernici ad acqua.

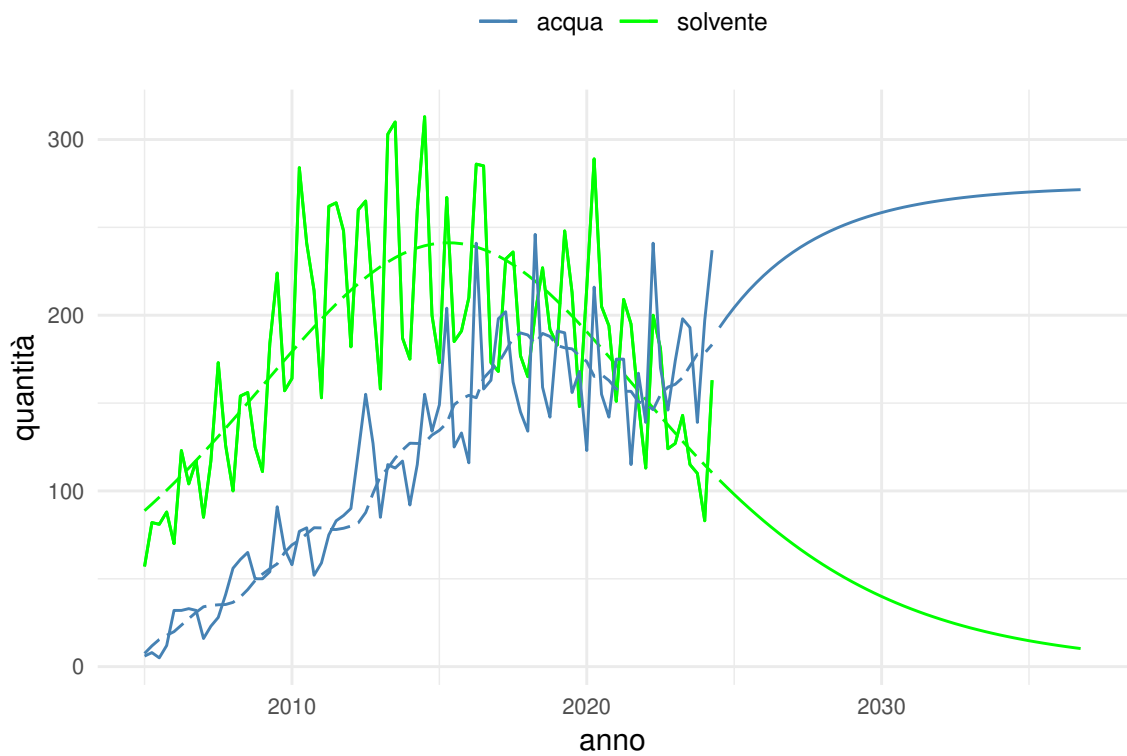


Figura 4.4: previsioni di vendita a lungo termine di smalti a solvente e ad acqua calcolate con i modelli 4.1 e 4.2.

2. Scenario di previsione a medio-breve termine

In questo secondo scenario viene assunto, invece, che la normativa rimanga, almeno per alcuni anni, ancora quella attuale. Questo comporterebbe che il passaggio dalle vernici a solvente a quelle ad acqua avvenisse più gradualmente, in modo simile a come è successo nell'ultimo decennio.

Per la serie delle vendite di vernici a solvente è stato scelto un modello SARIMA(1,1,1)(1,0,1)₁₂. L'equazione del modello è la seguente:

$$(1 - \phi_1 B - \Phi_1 B^{12})(1 - B)y_t = (1 - \theta_1 B - \Theta_1 B^{12})\epsilon_t \quad (4.3)$$

	AR1	MA1	SAR1	SMA1
coefficiente	0.2159	-0.9446	0.9471	-0.7358
errore standard	0.0831	0.0258	0.0393	0.0983

Tabella 4.3: stima dei parametri del modello 4.3.

Per la serie delle vernici a base acqua, invece, è stato scelto un modello SARIMA(1,0,1)(1,1,1)₁₂ con drift, ovvero:

$$(1 - \phi_1 B - \Phi_1 B^{12})(1 - B^{12})y_t = (1 - \theta_1 B - \Theta_1 B^{12})\epsilon_t + c \quad (4.4)$$

	AR1	MA1	SAR1	SMA1	drift
coefficiente	0.9687	-0.8524	-0.3658	-0.5969	0.2705
errore standard	0.0294	0.0589	0.1204	0.1551	0.1174

Tabella 4.4: stima dei parametri del modello 4.4.

Test di Ljung-Box		
	modello 4.3	modello 4.4
statistica	25.68969	24.72228
gdl	23	23
p-value	0.3157324	0.3647619

Tabella 4.5: Risultati del test di Ljung-Box per testare l'incorrelazione dei residui per il modello SARIMA.

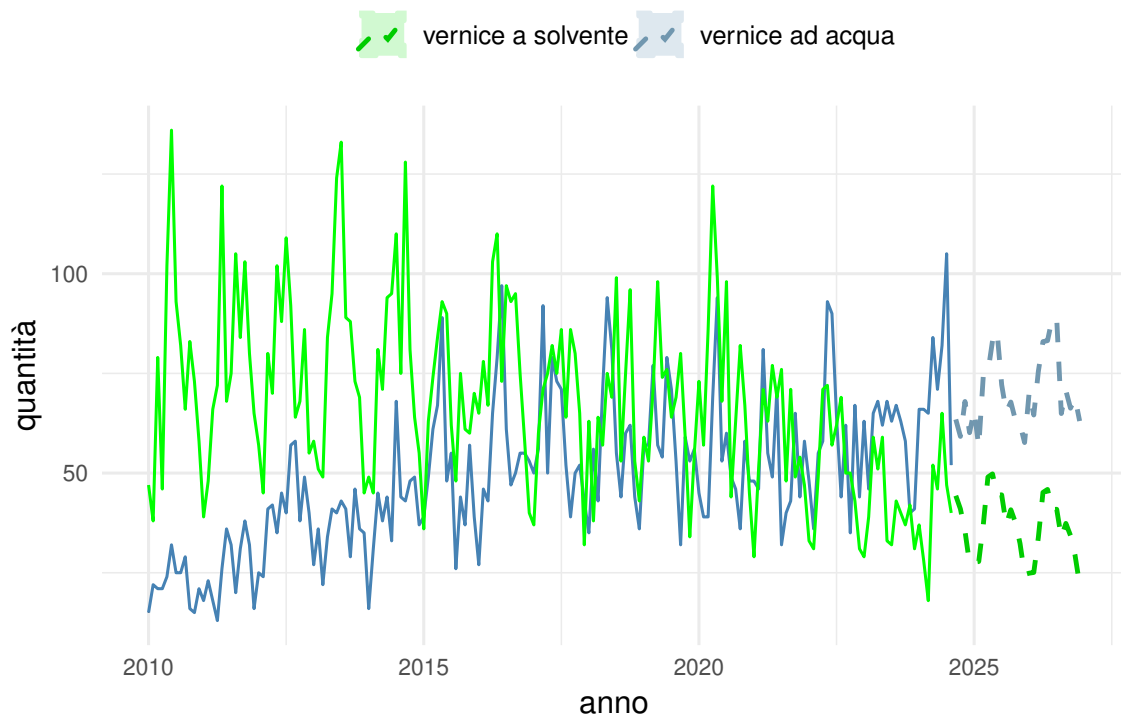


Figura 4.5: previsioni di vendita di smalti a solvente e ad acqua calcolate con i modelli SARIMA 4.3 e 4.4.

Capitolo 5

Conclusioni

Tutte le decisioni strategiche aziendali, in un modo o nell'altro, implicano previsioni sul futuro. Per affrontare al meglio alcune delle principali problematiche di un'attività commerciale, sono stati proposti alcuni approcci statistici, necessari per ottenere previsioni affidabili e di qualità. In particolare, l'attenzione posta alla gestione della stagionalità ha consentito di adattare modelli in grado di fornire risultati utili per l'ottimizzazione delle scorte e la gestione del magazzino. L'analisi del rischio di obsolescenza, invece, ha evidenziato come l'adozione di modelli di diffusione possa supportare efficacemente la gestione dei prodotti in settori caratterizzati da rapidi cambiamenti tecnologici. Infatti, per articoli come le chiavi, che possono diventare obsoleti in seguito all'evoluzione di nuove tecnologie, un processo previsionale accurato può permettere di individuare le fasi del ciclo di vita del prodotto, consentendo all'azienda di aggiornarsi e mantenersi competitiva rispetto alle novità del mercato. Infine, l'analisi delle tendenze di mercato, applicata a prodotti come le vernici, ha messo in luce alcune dinamiche evolutive proprie del settore, evidenziando opportunità di crescita per i prodotti ecocompatibili e una tendenza calante per i prodotti più tradizionali.

In sintesi, l'integrazione di modelli di previsione nei processi aziendali può risultare una scelta strategica valida, capace di dare all'azienda strumenti e risultati utili per prendere decisioni e rispondere prontamente ai cambiamenti nel mercato.

Appendice A

Codice R

In questa appendice verranno elencati i codici di R utilizzati per stimare i modelli proposti nei vari capitoli. Di seguito sono riportate le due librerie principali per la stima dei modelli.

```
# stima dei modelli, eccetto quelli di diffusione
library(forecast)
# stima dei modelli di diffusione
library(DIMORA)
```

A.1 Modelli del capitolo 1

Serie trimestrale

```
### modello di regressione con variabili dummy

# variabile dummy per includere l'effetto della
# pandemia
covid <- c(rep(0, 21), rep(1, 4), rep(0, 13))

# stima del modello
fit1.reg <- tslm(pennelli ~ trend + season + covid)
summary(fit1.reg)

# test di Durbin-Watson
dwtest(fit1.reg)
```

```
# previsioni
covid.previsioni <- c(rep(0, 12))
previsioni <- forecast(fit1.reg, data.frame(covid=covid.
  previsioni))

# analisi grafica dei residui
residui <- residuals(fit1.reg)
ggtsdisplay(residui, plot.type = 'histogram')
```

Serie mensile

```
### modello di Holt-Winters

# stima del modello con componente stagionale
# moltiplicativa
mod1 <- hw(pennelli.mensile, seasonal = 'multiplicative',
  damped=T, h=24)
summary(mod1)

# valori progressivi delle componenti del modello
mod1$model$states

### modello SARIMA

# funzione di autocorrelazione e autocorrelazione parziale
ggAcf(pennelli.mensile)
ggPacf(pennelli.mensile)

# differenziazione stagionale
dif12 <- diff(pennelli.mensile, lag=12)

# stima del modello
a1 <- Arima(pennelli.mensile, order=c(3,0,0), seasonal = c
  (0,1,1), include.drift = T)
summary(a1)

# analisi grafica dei residui
a1.res <- residuals(a1)
ggtsdisplay(a1.res, plot.type = 'histogram')
```

```
# test di Ljung-Box
Ljung.Box(a1.res, lag = 24)

### modello SARIMAX

# variabile dummy
covid2 <- ts(c(rep(0, 63), rep(1, 12), rep(0, 41)), start
  = 2015, frequency = 12)

# stima del modello con auto.arima
sarimax <- auto.arima(pennelli.mensile, xreg = covid2)
summary(sarimax)

# test di Ljung-Box
Ljung.Box(residuals(sarimax), lag=24)

### scelta del modello

# definizione dell'intervallo per il campione di stima e
# per il campione di test
training <- window(pennelli.mensile, end=c(2023, 8))
test <- window(pennelli.mensile, start=c(2023, 9))

# modelli adattati sul campione di stima
train_hw <- hw(training, seasonal = 'multiplicative',
  damped = T, h=12)
train_sarima <- Arima(training, order=c(3,0,0), seasonal =
  c(0,1,1), include.drift = T)
train_sarimax <- Arima(training, order=c(0,0,3), seasonal
  = c(0,1,1), include.drift = T, xreg = window(covid2,
  end=c(2023, 8)))

# previsioni sui dati disponibili
for_hw <- forecast(train_hw, h=12)
for_sarima <- forecast(train_sarima, h=12)
for_sarimax <- forecast(train_sarimax, xreg = window(covid2
  , start=c(2023,9)), h=12)

# misure di accuratezza
round(accuracy(for_hw, test), 3)
round(accuracy(for_sarima, test), 3)
round(accuracy(for_sarimax, test), 3)
```

A.2 Modelli del capitolo 2

```

### modello di Bass generalizzato con shock esponenziale

# stima del modello
bass_exp <- GBM(chiavi.tot, shock = "exp", nshock = 1,
  prelestimates = c(BM(chiavi.tot, display=FALSE)$
    Estimate[1,1], BM(chiavi.tot, display=FALSE)$
    Estimate[2,1], BM(chiavi.tot, display=FALSE)$
    Estimate[3,1], 24, 0.1, 0.5), oos=0)

# calcolo delle stime e delle previsioni istantanee
stime_exp <- ts(make.instantaneous(bass_exp$fitted), start
  = 2010, frequency = 4)
prev_exp <- make.instantaneous(predict.Dimora(bass_exp,
  newx = c(58:180)))
prev_exp <- ts(prev_exp[2:100], start = c(2024, 3),
  frequency = 4)

### modello di Bass con errori SARIMA

## stima del modello di Bass Standard
bass <- BM(chiavi.tot, oos = 50)
stime.bass <- ts(make.instantaneous(bass$fitted),
  start = 2010, freq=4)

# analisi grafica dei residui
res.bas <- residuals(bass)
ggtsdisplay(res.bas, plot.type = 'histogram')

# test di Durbin-watson
tempo <- 1:58
dwtest(res.bas ~ tempo, alternative = 'two.sided')

## stima del modello SARIMAX
ggtsdisplay(chiavi.tot)
diff4 <- diff(chiavi.tot, lag=4)
ggtsdisplay(diff4)
sarimax <- Arima(chiavi.tot, order = c(3,0,0),
  seasonal = c(0,1,1), xreg = stime.bass)
summary(sarimax)

```



```
# calcolo delle previsioni
prev.bass <- make.instantaneous(predict.Dimora(bass, newx
  = c(58:140)))[2:24]
prev.arimax <- forecast(sarimax, xreg = prev.bass)

# analisi grafica dei residui
ggtsdisplay(sarimax$residuals, plot.type = 'histogram')
```

A.3 Modelli del capitolo 3

```
### scenario a lungo termine

## stima del modello di Bass per le vernici a solvente
bass.solvente <- BM(solvente.tot, oos=50)
summary(bass.solvente)
stime.bass.solvente <- ts(make.instantaneous(bass.solvente$
  fitted), start = 2005, freq=4)

# calcolo delle previsioni
prev.bass.solvente <- make.instantaneous(predict.Dimora(
  bass.solvente, newx = c(78:140)))[2:51]

## stima del modello di Holt smorzato per le vernici ad
  acqua
fit <- holt(acqua.tot, damped = T, h=50)
summary(fit)

### scenario a medio-breve termine

## stima del modello SARIMA per le vernici a solvente
tsdisplay(solvente.tot)
diff1 <- diff(solvente.tot, lag=1)
ggtsdisplay(diff1)
mod1 <- Arima(solvente.tot, order = c(1,1,1),
  seasonal = c(1,0,1), include.drift = F)
summary(mod1)

# test di Ljung-Box
Ljung.Box(residuals(mod1), lag = 24)
```

```
# analisi grafica dei residui
ggtsdisplay(residuals(mod1))

## stima del modello SARIMA per le vernici ad acqua
tsdisplay(acqua.tot)
diff1_2 <- diff(acqua.tot, lag=1)
tsdisplay(diff1_2)
mod2 <- Arima(acqua.tot, order = c(1,0,1),
              seasonal = c(1,1,1), include.drift = T)
summary(mod2)

# test di Ljung-Box
Ljung.Box(residuals(mod2), lag=24)

# analisi grafica dei residui
ggtsdisplay(residuals(mod2))
```

Appendice B

Modelli di previsione

B.1 Modello di regressione per la stima di trend e stagionalità

I modelli di regressione possono essere utilizzati in modo efficace per fini previsivi nelle serie storiche economiche. In particolare, se nella serie sono presenti sia trend che stagionalità, la stima delle componenti può avvenire in maniera simultanea. La stagionalità viene modellata attraverso variabili dummy, mentre il trend viene descritto da una funzione polinomiale di grado q (Di Fonzo, T. e Lisi, F. 2005). Il modello di regressione in forma matriciale è così definito:

$$y = \tilde{P}\tilde{\alpha} + D\gamma + \epsilon, \quad (\text{B.1})$$

dove:

$$\tilde{P} = \begin{bmatrix} 1 & 1^2 & \dots & 1^q \\ 2 & 2^2 & \dots & 2^q \\ \dots & \dots & \dots & \dots \\ n & n^2 & \dots & n^q \end{bmatrix}; \quad \tilde{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_q \end{bmatrix}$$

Le stime di $\tilde{\alpha}$ e γ sono date da:

$$\begin{bmatrix} \hat{\tilde{\alpha}} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \tilde{P}'\tilde{P} & \tilde{P}'D \\ D'\tilde{P} & D'D \end{bmatrix}^{-1} \begin{bmatrix} \tilde{P}'y \\ D'y \end{bmatrix} \quad (\text{B.2})$$

B.2 Modelli di lisciamiento esponenziale

B.2.1 Modello di Holt con trend smorzato

Il modello di Holt è un'estensione del modello di lisciamiento esponenziale semplice in grado di fare previsione su dati che presentano una componente di trend. Nel modello di lisciamiento esponenziale di Holt smorzato l'effetto del trend viene ridotto tramite il parametro ϕ . In questa versione, la stima delle componenti del modello è descritta dalle seguenti equazioni di lisciamiento:

$$\begin{aligned}\hat{l}_t &= \alpha Y_t + (1 - \alpha)(\hat{l}_{t-1} + \phi \hat{b}_{t-1}) \\ \hat{b}_t &= \beta(\hat{l}_t - \hat{l}_{t-1}) + (1 - \beta)\phi \hat{b}_{t-1},\end{aligned}$$

dove \hat{l}_t è la stima del livello della serie al tempo t e \hat{b}_t la stima del trend (pendenza) della serie sempre al tempo t . α è il parametro di lisciamiento per il livello e β il parametro di lisciamiento per il trend. Tutti i parametri sono compresi tra 0 ed 1. Nel contesto di stima, è consuetudine impiegare dei parametri per cui viene minimizzata la somma dei quadrati dei residui, ovvero:

$$SSE = \sum_{t=1}^T (y_t - \hat{y})^2 = \sum_{t=1}^T e_t^2$$

Lo stesso vale per la scelta dei valori di inizializzazione del livello l_0 e del trend b_0 .

Le previsioni sono date da:

$$\hat{y}_{t+m} = \hat{l}_t + (\phi + \phi^2 + \dots + \phi^m)\hat{b}_t. \quad (\text{B.3})$$

La funzione di previsione, dunque, non è piatta come nel lisciamiento esponenziale semplice, bensì tendenziale. Per qualsiasi ϕ tra 0 ed 1, le previsioni di breve periodo sono dotate di trend, mentre quelle di lungo periodo sono costanti. Infatti, le previsioni convergono a $l_t + \frac{\phi b_t}{1-\phi}$ per $m \rightarrow \infty$. Nella pratica ϕ è raramente inferiore a 0.8 poiché lo

smorzamento ha un effetto molto forte per valori più piccoli. Questa formulazione consente di migliorare l'accuratezza delle previsioni, rendendole più realistiche a lungo termine.

B.2.2 Modello di Holt-Winters con trend smorzato e stagionalità moltiplicativa

Nel modello di lisciamiento esponenziale di Holt-Winters con stagionalità additiva la componente stagionale è espressa in termini relativi. All'interno di ogni anno, la componente stagionale sommerà circa ad s , ossia il periodo della serie. La stima delle componenti del modello con trend smorzato è data dal seguente sistema di equazioni:

$$\begin{aligned}\hat{l}_t &= \alpha \frac{Y_t}{\hat{S}_{t-s}} + (1 - \alpha)(\hat{l}_{t-1} + \phi \hat{b}_{t-1}) \\ \hat{b}_t &= \beta(\hat{l}_t - \hat{l}_{t-1}) + (1 - \beta)\phi \hat{b}_{t-1} \\ \hat{S}_t &= \gamma \frac{Y_t}{(\hat{l}_{t-1} + \phi \hat{b}_{t-1})} + (1 - \gamma)\hat{S}_{t-s}\end{aligned}$$

Le previsioni sono date da:

$$\hat{y}_{t+m} = [\hat{l}_t + (\phi + \phi^2 + \dots + \phi^m)\hat{b}_t m] \hat{S}_{t-s+m}, \quad (\text{B.4})$$

B.3 Modelli *SARIMA*

I processi stocastici ARIMA stagionali, comunemente indicati come SARIMA, sono un'estensione ideata da Box e Jenkins per permettere ai modelli ARIMA di catturare comportamenti di tipo periodico. Un modello SARIMA(p, d, q)(P, D, Q) $_s$ è definito da:

$$\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D Y_t = c + \theta(B)\Theta(B^s)\epsilon_t, \quad (\text{B.5})$$

dove:

- s è il periodo stagionale
- $\phi(B)$ è l'operatore autoregressivo non stagionale di ordine p .
- $\Phi(B^s)$ è l'operatore autoregressivo stagionale di ordine P .
- $\theta(B)$ è l'operatore a media mobile non stagionale di ordine q .
- $\Theta(B^s)$ è l'operatore a media mobile stagionale di ordine Q .
- $(1 - B)^d$ è l'operatore differenza non stagionale di ordine d .
- $(1 - B)^D$ è l'operatore differenza stagionale di ordine D .
- ε_t è un processo *White Noise*.

La stima di un modello SARIMA può essere sviluppata attraverso la procedura di Box e Jenkins, la quale prevede tre fasi: identificazione, stima dei parametri e controllo diagnostico.

Identificazione

L'identificazione consiste nel riconoscere nella struttura della funzione di autocorrelazione della serie storica la struttura di una ACF teorica.

La componente stagionale autoregressiva o a media mobile di un modello si può vedere ai ritardi stagionali della PACF e ACF.

L'assunzione fondamentale alla base di questo procedimento è che la serie sia stazionaria, ovvero caratterizzata da media e varianza costanti nel tempo. Nel caso in cui la serie di partenza mostri un andamento non stazionario, è necessario differenziarla (non stazionarietà in media) o applicarle una trasformazione logaritmica (non stazionarietà in varianza).

Stima dei parametri

Dopo aver identificato gli ordini del modello, ossia i valori di p, d, q, P, D, Q , a partire dalla ACF e dalla PACF della serie occorre stimare i parametri associati $(\phi_0, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2)$. Uno dei

metodi più utilizzati è quello della stima di massima verosimiglianza. Questa tecnica trova i valori dei parametri che massimizzano la probabilità di ottenere i dati che sono stati osservati. Per i modelli SARIMA, il metodo della massima verosimiglianza è simile al metodo di stima dei minimi quadrati che minimizza la quantità:

$$\sum_{t=1}^n \epsilon_t^2$$

B.4 Modello di regressione dinamica SARIMAX

Un modello di regressione con errori ARIMA stagionali, detto anche SARIMAX, è definito dalla seguente formulazione:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + \eta_t,$$

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D\eta_t = \theta(B)\Theta(B^s)\epsilon_t. \quad (\text{B.6})$$

Il modello ha quindi due termini di errore; l'errore del modello di regressione, indicato con η_t , e l'errore del modello SARIMA, indicato con ϵ_t . Gli errori del modello SARIMA sono assunti white noise. La stima dei parametri del modello si ottiene minimizzando la somma dei valori al quadrato di ϵ_t , e non quella degli errori η_t . Quando si stima una regressione con errori SARIMA è importante assicurarsi che tutte le variabili nel modello siano stazionarie. La stima dei parametri può essere calcolata anche attraverso il metodo della massima verosimiglianza. Per fare previsioni utilizzando un modello di regressione con errori SARIMA, si deve prevedere la parte di regressione del modello e la parte SARIMA del modello, e combinare i risultati. Come per i modelli di regressione classici, per ottenere le previsioni è necessario prima avere i valori futuri dei predittori. Quando i predittori sono a loro volta non noti, bisogna modellarli separatamente, oppure usare dei valori futuri presunti.

B.5 Modelli di diffusione

B.5.1 Modello di Bass standard

Il Modello di Bass (1969) è caratterizzato da un'equazione differenziale che descrive il ciclo di vita delle innovazioni o dei prodotti in un sistema economico. Il modello si basa sull'idea che il processo di adozione di un'innovazione sia influenzato da due fonti di informazione, una esterna ed una interna. La prima comprende la pubblicità e i canali di comunicazione; la seconda, invece, è l'imitazione. Queste due fonti determinano rispettivamente la formazione di due gruppi di utilizzatori, gli innovatori e gli imitatori. L'equazione differenziale è la seguente:

$$z'(t) = \left(p + q \frac{z(t)}{m} \right) (m - z(t)), \quad t > 0 \quad (\text{B.7})$$

dove:

- m è il mercato potenziale,
- p è il coefficiente di innovazione,
- q è il coefficiente di imitazione,

Nell'equazione B.7, la variazione nel tempo delle adozioni $z'(t)$ è proporzionale al mercato residuo $m - z(t)$, dove $z(t)$ è numero di adozioni cumulato al tempo t . Il mercato potenziale m viene assunto costante durante l'intero processo di diffusione. I parametri p e q misurano i due diversi tipi di consumatori (innovatori e imitatori). L'equazione differenziale B.7 ha una forma chiusa, che è data da:

$$z(t) = m \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}}$$

B.5.2 Modello di Bass generalizzato con shock esponenziale

Il modello di Bass generalizzato contiene nella sua struttura una funzione $x(t)$ di intervento che permette di modificare l'adattamento del modello in intervalli opportuni. Il modello di Bass generalizzato è descritto da:

$$z'(t) = \left(p + q \frac{z(t)}{m} \right) (m - z(t)) x(t), \quad t > 0 \quad (\text{B.8})$$

con $x(t)$ funzione integrabile in domini limitati e non negativa. Nel caso particolare del modello di Bass questa funzione assume valore uguale ad 1. Il valore di $x(t)$ può determinare il rallentamento o l'accelerazione della diffusione. La soluzione in forma chiusa del modello di Bass generalizzato (equazione B.8) è data da:

$$z(t) = m \frac{1 - e^{-(p+q) \int_0^t x(\tau) d\tau}}{1 + \frac{q}{p} e^{-(p+q) \int_0^t x(\tau) d\tau}}. \quad (\text{B.9})$$

La funzione $x(t)$, dunque, non altera il valore del mercato potenziale, bensì il processo di diffusione dell'innovazione. Tale funzione può assumere qualsiasi valore entro le condizioni di integrabilità e non negatività. Nel contesto delle vendite, può essere utile descrivere perturbazioni intense e molto brevi. In questi casi si possono integrare degli impulsi esponenziali. Uno shock esponenziale è descritto come:

$$x(t) = 1 + c_1 e^{b_1(t-a_1)} I_{t \geq a_1} \quad (\text{B.10})$$

con I_A funzione indicatrice che assume valore 1 se A è vero e 0 altrimenti. Dove:

- a_1 rappresenta il tempo d'insorgenza dello shock
- b_1 misura la velocità di decadimento dello shock
- c_1 è l'intensità dello shock (valore iniziale), che può essere sia positiva che negativa.

Appendice C

Test statistici e criteri per il confronto tra modelli

C.1 Test per l'incorrelazione dei residui

C.1.1 Test di Durbin-Watson

Nei residui delle serie storiche spesso è presente autocorrelazione seriale di primo ordine. Ad esempio, in un modello di regressione lineare multipla con componente di errore ϵ_t , l'autocorrelazione di ordine uno è così definita:

$$\epsilon_t = \rho\epsilon_{t-1} + \eta_t, \quad (\text{C.1})$$

con ρ indice di correlazione e η_t componente erratica di media nulla e varianza costante. Se $\rho = 0$, allora $\epsilon_t = \eta_t$.

Uno dei test più utilizzati per controllare questo tipo di autocorrelazione è il test di Durbin-Watson, il quale è definito dalla seguente formulazione di ipotesi:

$$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho > 0 \end{cases}$$

L'ipotesi alternativa, in questo caso, non è $\rho \neq 0$ perchè nella maggior parte dei casi le serie storiche economiche sono caratterizzate da autocorrelazione positiva. La statistica di Durbin-Watson è data da:

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (C.2)$$

Per grandi campioni vale l'equazione:

$$DW = 2(1 - r_1(e)), \quad (C.3)$$

dove $r_1(e)$ è l'autocorrelazione di primo ordine nei residui. La statistica di Durbin-Watson assume valori tra 0 e 4, con $DW = 2$ valore centrale.

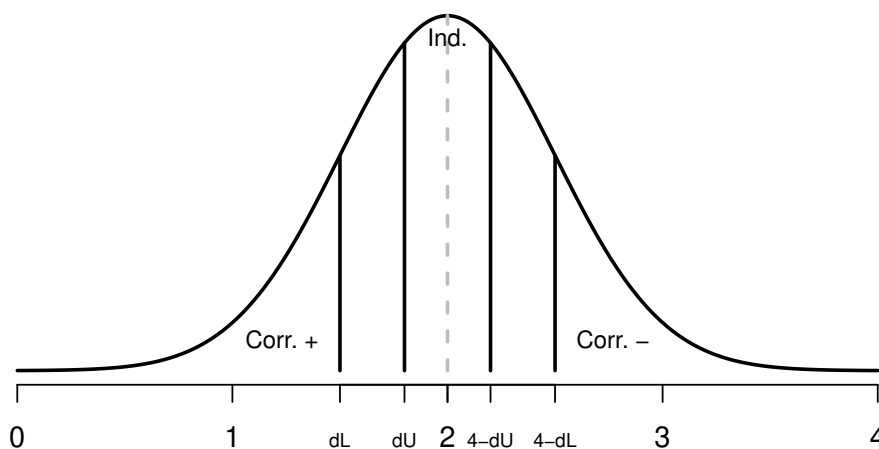


Figura C.1: distribuzione della statistica di Durbin-Watson.

Se $DW > d_u$ allora si accetta l'ipotesi nulla, altrimenti occorre controllare la forma funzionale del modello. In alcuni casi potrebbe essere necessario aggiungere una variabile esogena in grado di spiegare la dipendenza temporale nella serie.

C.1.2 Test di Ljung-Box

Per una verifica globale dell'assenza di correlazione nei residui si può utilizzare la statistica di Ljung-Box, la quale si basa sul seguente sistema di ipotesi:

$$\begin{cases} H_0 : \rho_1 = \rho_2 = \dots = \rho_q = 0 \\ H_1 : \rho_k \neq 0 \text{ per almeno un } k \in \{1, 2, \dots, m\} \end{cases}$$

La statistica test del test di Ljung-Box (Ljung, G.M. e Box, G.E.P. 1978) è data dalla formula:

$$Q(m) = n(n+2) \sum_{k=1}^m \frac{1}{n-k} \hat{\rho}_k^2 \sim \chi^2(m-p),$$

dove m è il numero di ritardi presi in considerazione e $\hat{\rho}_k^2$ è l'autocorrelazione campionaria al lag k .

C.2 Criteri per la selezione del modello

C.2.1 R^2 e R^2 corretto

Nel caso dei modelli di regressione è utile poter selezionare il miglior insieme di predittori da usare. L'output di un modello di regressione comprende sempre il valore dell' R^2 . Esso può essere calcolato nel seguente modo:

$$R^2 = \frac{\sum (\hat{y}_t - \bar{y}_t)^2}{\sum (y_t - \bar{y}_t)^2} \quad (\text{C.4})$$

Il coefficiente di determinazione R^2 indica quanto bene un modello di regressione lineare si adatta ai dati. Tale misura riflette la proporzione di varianza spiegata dal modello. Tuttavia, questo indicatore non rappresenta una buona misura dell'abilità previsiva di un modello. Includendo una qualsiasi variabile, si ottiene un incremento del valore dell' R^2 , anche se la variabile inclusa non è significativa. Una strategia

più efficace per valutare un modello di regressione consiste nell'impiego dell' R^2 corretto:

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - k - 1}, \quad (\text{C.5})$$

con k numero di predittori.

C.2.2 Criteri di informazione

I criteri di informazione sono misure utili per la selezione dei modelli statistici. I più utilizzati sono l'AIC (criterio di informazione di Akaike) e il BIC. Entrambi cercano di bilanciare la bontà di adattamento del modello con la sua complessità, penalizzando l'aggiunta di parametri. Sia per l'AIC che per il BIC, valori più bassi indicano un modello migliore.

Il criterio di informazione di Akaike

Il criterio di informazione di Akaike (Akaike 1974) è dato da:

$$\text{AIC} = 2k - 2 \ln(\ell(\hat{\theta})),$$

in cui k è il numero di parametri del modello e $\ell(\hat{\theta})$ è il valore della funzione di log-verosimiglianza del modello in $\hat{\theta}$, ovvero il vettore $(k \times 1)$ contenente i parametri stimati.

Criterio di informazione bayesiano di Schwarz

Il criterio di informazione bayesiano (Schwarz 1978) è dato da:

$$\text{BIC} = k \ln(n) - 2 \ln(\ell(\hat{\theta}))$$

Il BIC tende a selezionare modelli più semplici rispetto all'AIC, specialmente quando n è elevato, poiché la penalizzazione aumenta con il logaritmo delle osservazioni.

Bibliografia

- Akaike, H. (1974). “A new look at the statistical model identification”.
In: *IEEE Transactions on Automatic Control* 19.6, pp. 716–723.
- Di Fonzo, T. e Lisi, F. (2005). *Serie storiche economiche: Analisi statistiche e applicazioni*. Carocci editore. ISBN: 9788843034239.
- Guidolin, M. (2023). *Innovation Diffusion Models: Theory and Practice*. Wiley. ISBN: 9781119756200.
- Hyndman, R.J. e Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts. ISBN: 9780987507112.
- Ljung, G.M. e Box, G.E.P. (1978). “On a measure of a lack of fit in time series models”. In: *Biometrika* 65.2, pp. 297–303.
- Paci, A. (2019). *Fondamenti di strategia. Il framework di Exploring Strategy*. Pearson. ISBN: 9788891911056.
- Schwarz, G.E. (1978). “Estimating the dimension of a model”. In: *Annals of Statistics* 6.2, pp. 461–464.

Sitografia

Assofermet (2023). *Ferramenta: 17 miliardi di fatturato per il settore.*

URL: <https://www.assofermet.it/comunicati/ferramenta-17-miliardi-di-fatturato-per-il-settore-34-6-fra-2019-e-2022-pi-ugrave-di-72-mila-dipendenti?cookie=1>.

Unione Europea (2004). *Emissioni di composti organici volatili in*

pitture, vernici e prodotti per carrozzeria. URL: <https://eur-lex.europa.eu/IT/legal-content/summary/emissions-of-volatile-organic-compounds-in-paints-varnishes-and-vehicle-refinishing-products.html>.