

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche



**“Previsione della domanda commerciale
promozionale nella GDO”**

Relatore Dott. Mariangela Guidolin
Dipartimento di Scienze Statistiche

Laureando: Daniel Gazzea
Matricola N 1132262

Anno Accademico 2017/2018

Indice

Introduzione	5
1. Time Series Analysis	9
1.1 Obiettivi della time series analysis	11
1.2 Componenti delle serie storiche.....	11
1.3 Tipologie di serie storiche.....	13
2. Sales cleaning	17
2.1 Metodo della scomposizione moltiplicativa.....	19
2.2 Metodo della scomposizione additiva.....	22
2.3 Metodo dell'IC per la rettifica degli outliers.....	24
2.4 Metodo della Media mobile per la rettifica degli outliers.....	26
2.5 Metodo dell'impatto promozionale.....	29
3. Valutazione dei modelli di Sales Forecasting	33
3.1 Training e test set standard	34
3.2 One-step-ahead forecasts	35
3.3 One-step-promotional-ahead forecasts	36
3.4 Misure di accuratezza previsionale.....	37
4. Times Series Forecasting	41
4.1 Fasi del processo di sales forecasting.....	41
4.2 Classificazione dei modelli di time Series Forecasting.....	44
4.3 Metodo informale: Naïve Forecasting.....	45
4.4 Metodo delle serie storiche.....	45
4.4.1 Modello di liscio esponenziale semplice.....	47
4.4.2 Modello di Holt-Winters.....	48
4.4.3 Modello di Croston.....	52
4.4.4 Processi stocastici: WN,AR,ARMA,ARIMA,SARIMA.....	54
4.5 Modelli di regressione.....	64
4.5.1 Modello di regressione lineare.....	66
4.5.2 Modello di reg. lineare con trasformata logaritmica....	68
4.6 Modelli di previsione avanzati: Rete Neurale.....	69
4.7 Modelli per il new product forecasting.....	71
4.8 Combinazione di modelli di sales forecasting.....	73

5. Il caso di studio	75
5.1 Fasi del processo.....	76
5.2 Analisi preliminare e codifica variabili.....	79
5.3 Sales cleaning e sales forecasting applicati ai casi di studio.....	80
5.4 Modelli di sales forecasting nei casi di studio	84
5.5 Caso di studio 1.....	87
5.6 Caso di studio 2.....	95
5.7 Caso di studio 3.....	102
6. Conclusione	111
7. Riferimenti bibliografici	113

Introduzione

Nel mondo della Grande Distribuzione Organizzata (GDO) affinché i punti vendita siano efficienti è necessario fare affidamento sulla previsione della domanda commerciale dei prodotti, che in questo contesto vengono chiamati SKU (dall'inglese *Stock Keeping Units*). Infatti grazie a queste informazioni, e alla conoscenza relativa alla gestione delle scorte, i punti vendita riescono a ordinare un quantitativo di merce, necessario per coprire un orizzonte temporale pianificato (lo stesso usato nella generazione della previsione) come ad esempio un giorno o una settimana, in modo che non sia né troppo elevato e quindi evitare elevati costi di inventario con relativo intasamento del magazzino, né troppo basso per non incorrere in problemi dovuti alla rottura di stock e quindi alla perdita di vendite, e potenzialmente anche degli stessi clienti.

Per queste ragioni la previsione della domanda commerciale (*Demand Planning, Sales Forecasting*) relativa al comparto della GDO rappresenta un processo ciclico di *Supply Chain Planning* che assume importanza strategica per le imprese industriali e commerciali operanti in questo settore merceologico (Dallari e Creazza, 2006).

Nell'oceano rosso del settore della GDO, un importante elemento per riuscire a sopravvivere alla sempre maggiore competizione è rappresentato dalle campagne promozionali, che servono alle aziende per mantenere i clienti già esistenti e attrarre nuovi consumatori sottraendoli ai competitor.

Tuttavia, se da un lato la previsione della domanda commerciale per prodotti che non sono in promozione non presenta particolari problemi, in quanto possono essere utilizzate tecniche di *sales forecasting* tradizionali (e.g. Makridakis et al., 1998), dall'altro le previsioni per prodotti soggetti ad attività promozionali risultano molto difficili.

Infatti i prodotti in promozione lo sono tipicamente solo per un limitato periodo di tempo, ad esempio una settimana, durante la quale la domanda è solitamente molto più alta rispetto ai periodi senza promozioni.

In letteratura sono stati condotti molti studi sull'effetto delle promozioni sulle vendite considerando rilevanti i seguenti fattori: la variazione tra prezzo regolare e prezzo promozionale per un particolare SKU (Blattberg et al., 1995; Christen et al., 1997; Lattin and Bucklin, 1989; Mulhern et al., 1991; Cooper et al., 2000); la frequenza delle promozioni per prodotti simili (Christen et al., 1997); il metodo pubblicitario utilizzato, ad esempio giornali, tv, radio, etc. (e.g. Baltas, 2005; Narasimhan et al., 1996; Blatterg et al., 1995); attività promozionali dei competitor (Struse, 1987), precedenti promozioni per lo stesso prodotto e numero di varianti del prodotto in promozione. Questi ultimi

due aspetti sono legati al cosiddetto problema della cannibalizzazione della domanda da parte della stessa organizzazione (vedasi e.g. Neslin, 1990).

Tuttavia tutti questi studi non hanno come obiettivo di prevedere la domanda commerciale promozionale, ma solo di determinare i principali effetti che contribuiscono all'aumento della domanda stessa.

In questo elaborato si vogliono ampliare le tecniche di *sales forecasting* tradizionali per prevedere la domanda commerciale regolare, adattandole alla previsione della domanda commerciale per periodi promozionali.

Qui di seguito si riporta un breve riassunto di ognuno dei 5 capitoli in cui è suddiviso l'elaborato.

- Nel primo capitolo si descrivono le caratteristiche generali di una serie storica, e in particolare quelle di una serie di domanda, mostrando alcuni esempi. Inoltre si argomentano gli obiettivi della *time series analysis* orientati alla previsione della domanda.
- Nel secondo capitolo viene affrontato il problema del *sales cleaning*, ovvero quel processo necessario per ripulire la serie storica della domanda, in modo da renderla utilizzabile per i successivi processi di *sales forecasting*.
- Nel terzo capitolo viene discussa la valutazione dei modelli di *sales forecasting*, ovvero si vuole rispondere alla domanda su come individuare il migliore modello di sales forecasting da utilizzare per la previsione della domanda commerciale, sulla base di alcune misure di accuratezza previsionale (o alla loro combinazione) e grazie a 3 metodi specifici per calcolare il training set e il test set in base all'obiettivo dei *Demand Planner* (previsione domanda regolare o promozionale). Di questi solamente i primi due sono presenti in letteratura mentre il terzo viene implementato per la prima volta in questo elaborato proprio per risolvere il problema in esame ovvero quello della previsione promozionale.
- Nel quarto capitolo vengono descritte le fasi del processo di sales forecasting e i principali modelli di sales forecasting presenti in letteratura con le loro rispettive classificazioni. Particolare attenzione viene data ai modelli per il *new product forecasting*, ovvero quando la previsione non viene generata grazie alla serie storica di domanda del prodotto, perché quest'ultima non esiste o presenta poche osservazioni non sufficienti per produrre il forecast con i metodi standard. Viene inoltre discussa la possibilità di combinare i vari modelli

attraverso delle tecniche specifiche per migliorare il *best-fit* algoritmico.

- Nel quinto capitolo viene presentato il caso di studio orientato alla previsione della domanda commerciale promozionale utilizzando dati reali appartenenti a un'azienda della GDO italiana molto affermata nel Triveneto. Si farà particolare riferimento al processo dell'azienda e alla generazione del forecast sia utilizzando modelli di sales forecasting per serie storiche continue, sia modelli per il *new product forecasting*. Si prenderanno in considerazione solo SKU appartenenti alla promozione di volantino anche se la metodologia applicata può essere estesa alle altre tipologie promozionali aziendali.

1. Time series analysis

Una serie storica è un insieme di valori numerici ordinati in maniera sequenziale nel tempo osservati nel passato in corrispondenza di una variabile misurabile, come ad esempio la domanda commerciale o il prezzo (Di Fonzo e Lisi, 2013). In particolare nel seguito si farà riferimento solamente a *serie storiche economiche*. L'ordine delle osservazioni è fondamentale in questo contesto e non è totalmente casuale; infatti un'osservazione in un determinato tempo "t" ha una qualche forma di dipendenza solo ed esclusivamente dalle osservazioni precedenti, ma non da quelle successive (Wooldrige, 2006).

Le serie storiche, trattate nel seguito, possono essere pensate come delle realizzazioni di un processo stocastico, chiamato anche casuale, e non di un processo deterministico in quanto al tempo "t" è difficile conoscere con certezza il valore di un'osservazione futura, per esempio al tempo t+1. Per questa ragione le serie storiche, che vengono determinate da un processo stocastico come quelle trattate nel seguito, hanno le caratteristiche necessarie ad essere definite variabili casuali (Wooldrige, 2006).

Altre caratteristiche fondamentali delle serie storiche oggetto di studio, sono in primo luogo, il fatto che le osservazioni rilevate sono *equispaziate*, ciò significa che le osservazioni vengono raccolte regolarmente, sempre con una determinata frequenza che può essere giornaliera, settimanale, mensile, annuale; in secondo luogo, che le serie storiche siano univariate, ovvero in un determinato tempo "t" viene registrata un'unica osservazione del fenomeno di interesse.

Nel mondo della Grande Distribuzione Organizzata (GDO) le serie storiche, da qui in avanti chiamate serie storiche di domanda, vengono implementate grazie alla raccolta delle transazioni commerciali avvenute presso i punti vendita. Il metodo più semplice e veloce per registrare tali informazioni consiste nella scansione di un codice univoco ed identificativo del prodotto alle casse del negozio al momento della vendita dello stesso. Un sistema raccoglierà i dati di vendita giornalieri per ogni negozio e con una certa frequenza (di solito giornaliera e preferibilmente in orario notturno) li invierà al *datawarehouse* aziendale (DWH) che li storicizzerà al proprio interno. Il DWH potrà, quindi, essere interrogato per leggere le informazioni del venduto per quegli articoli e punti vendita oggetto di studio, con le quali vengono prodotte le previsioni della domanda (*sales forecast*) grazie ad algoritmi matematico-statistici.

In generale una serie storica di domanda viene indicata come $D_{t,p,c}$ dove t è l'indice temporale ($t = 1 \dots, T$), p si riferisce al prodotto e c al mercato (Milanato, 2008).

Ad esempio parlando di prodotto si può far riferimento al singolo SKU (Stock Keeping Unit), ovvero un articolo gestito a magazzino e messo a stock in una specifica posizione. In alternativa con prodotto si può intendere una famiglia o un particolare cluster di SKU caratterizzato da informazioni simili come, ad esempio, diversi gusti di uno yogurt. Invece, con il termine mercato ci si può riferire al singolo *store*, oppure a un raggruppamento di punti vendita per area geografica o dimensioni. Infine la dimensione temporale può essere giornaliera, settimanale, mensile o annuale.

I modelli previsionali (*sales forecasting*) analizzano le caratteristiche delle serie della domanda D_t proiettandole in un orizzonte temporale futuro per generare la previsione statistica $\hat{y}_t = F_t$ detta *sales forecast*. In generale una previsione è del tipo $F_t = f(D_t, D_{t-1}, D_{t-2}, \dots, D_{t-T})$

Per le analisi di *Demand Planning* è preferibile lavorare con dati storici presenti nel DW aggregati, sia nella dimensione temporale t , per esempio passando da giorno a settimana, sia a livello di mercato, per esempio passando dal singolo store a un gruppo di punti vendita accomunati dal fatto di vendere quello stesso SKU. Questo è utile per ottenere una previsione statisticamente più affidabile dal momento che “i dati aggregati lungo la dimensione tempo risultano più facilmente intelligibili ed interpretabili dagli algoritmi di scomposizione delle serie storiche, il cui obiettivo è la generazione di un sales forecast ottenuto da scomposizione e proiezione nel futuro delle componenti elementari della serie storica di partenza” (Milanato, 2008).

Da qui in avanti si parlerà di una serie di domanda $D_{t,p,c}$ con determinate caratteristiche, rispetto ad un'altra, in base all'obiettivo della previsione della domanda dei Demand Planner¹.

In particolare se l'obiettivo è quello di **prevedere la domanda commerciale regolare** si farà riferimento a una serie di domanda $D_{t,p,c}$ dove gli indici corrispondono a:

- t = settimane
- p = singolo SKU
- c = un raggruppamento di punti vendita per i quali la vendita dello SKU p è abilitata

Mentre se l'obiettivo dei Demand Planner è quello di **prevedere la domanda commerciale promozionale** si parlerà di una serie di domanda $D_{t,p,c}$ con le seguenti caratteristiche:

- t = settimane
- p = singolo SKU in offerta

¹ Persone con competenze statistico-matematiche addette alla previsione della domanda con software statistici all'interno dell'azienda

- c = un raggruppamento di punti vendita per cui quello SKU p andrà in promozione ovvero sarà venduto ad un prezzo inferiore al proprio prezzo regolare

Bisogna sottolineare il fatto che la metodologia descritta successivamente può essere applicata a qualunque altra combinazione di prodotto-mercato.

1.1 Obiettivi della Time Series Analysis

L'obiettivo dell'analisi delle serie storiche, o *time series analysis*, è quello di analizzare le caratteristiche quantitative delle serie di domanda D_t attraverso metodologie statistiche.

Un altro obiettivo riguarda l'eliminazione degli *outliers*, ovvero quei valori anomali, riscontrati nel passato non ad intervalli regolari che si discostano molto dal valore medio della serie. Questo è un passaggio fondamentale per l'accuratezza del forecast, in quanto un'attenta procedura di **sales cleaning** dai valori anomali rende la serie storica idonea ad essere studiata con i modelli di *sales forecasting*, altrimenti gli outliers potrebbero essere scambiati per fenomeni regolari e ricorrenti inficiandone la qualità predittiva (Milanato, 2008). Dalla serie storica originale D_t si procede rettificando i valori anomali e isolando le componenti regolari della serie storica che rappresentano il principale input per i modelli di *sales forecasting*, basati sulla scomposizione delle serie nelle componenti principali. Questi metodi statistici generano la previsione statistica della domanda, proiettando le componenti regolari della serie nel futuro e successivamente ricomponendole.

1.2 Componenti delle serie storiche

Le serie storiche possono essere distinte in due macroclassi in base alla presenza di valori pari a 0 all'interno delle stesse. In particolare si distinguono:

- *serie storiche continue*: dove la presenza di zeri è molto ridotta, quasi assente
- *serie storiche sporadiche*: dove esiste una percentuale rilevante di valori nulli, almeno pari al 20-30% del totale delle osservazioni della serie storica

Le ultime sono di difficile trattazione da parte degli algoritmi di *sales forecasting* poiché è molto difficile individuare componenti di regolarità e ciclicità all'interno di esse.

Le serie storiche intermittenti, particolare tipo di serie storica sporadica, sono caratterizzate da acquisti in modo non frequente, senza un'evidente regolarità temporale fra acquisti consecutivi, e in quantità molto variabili.

Nella seconda macroclasse per la distinzione di serie storiche, si individuano le serie storiche continue. Queste rappresentano la categoria maggiormente analizzata e per questo in letteratura esistono molti modelli di *sales forecasting* per la loro trattazione.

Le serie storiche continue sono caratterizzate da quattro componenti principali:

- Il *Trend* (o componente tendenziale) di una serie storica descrive la tendenza di fondo del fenomeno oggetto di studio, riferita ad un lungo periodo di tempo. Viene determinato prevalentemente da fenomeni quali lo sviluppo e l'evoluzione strutturale del sistema economico, che per loro natura si manifestano con gradualità e lentamente nel tempo. Per questo, si è soliti ipotizzare che i valori del trend siano esprimibili tramite una funzione sufficientemente regolare e liscia (*smooth*), ad esempio funzioni polinomiali o esponenziali. Se si assume che la funzione da utilizzare sia monotona, allora il trend può essere una funzione crescente o decrescente del fenomeno nel periodo analizzato (Di Fonzo e Lisi, 2013);
- la *Stagionalità* (o componente stagionale) "è costituita dai movimenti del fenomeno nel corso dell'anno che per effetto dell'influenza di fattori climatici e sociali, tendono a ripetersi in maniera pressoché analoga nel medesimo periodo (mese o trimestre) di anni successivi" (Di Fonzo e Lisi, 2013). Esempi di stagionalità possono essere gli aumenti dei consumi nelle ricorrenze di Natale o Pasqua;
- la *Ciclicità* (o componente congiunturale) di una serie storica "è costituita dalle fluttuazioni" ovvero movimenti alternati verso l'alto e verso il basso, "attribuibili al succedersi nel fenomeno considerato di fasi ascendenti e di fasi discendenti, generalmente collegate con le fasi di espansione e di contrazione dell'intero sistema economico" (Di Fonzo e Lisi, 2013). La differenza rispetto alla stagionalità, che dal punto di vista numerico è uguale alla ciclicità, è il fatto che la prima è nell'ordine degli anni per i cicli di lungo periodo, mentre la seconda dell'ordine dei mesi o al più un anno;
- la *Componente Casuale* (*rumore*) rappresenta il rumore di fondo della serie storica, cioè la componente di domanda non prevedibile, data dalla fluttuazione casuale dei valori di domanda attorno al valor medio della serie. La fluttuazione casuale viene rilevata dopo aver rimosso le tre componenti regolari della serie storica, avendo cioè isolato la

domanda media. Va evidenziato il fatto che la componente casuale non è statisticamente prevedibile; tuttavia se risulta essere numericamente rilevante, è possibile applicare modelli di regressione lineare nei quali vengono testate diverse variabili indipendenti sulla serie storica con la sola componente casuale. Questo è utile per scovare correlazioni tra componente casuale e variabili di input misurabili (per le quali è disponibile anche la previsione per i valori futuri).

1.3 Tipologie di serie storiche

Una classificazione per tipologia, il più possibile completa ed esaustiva, delle serie storiche è quella che si basa sull'eventuale presenza di una o più componenti regolari descritte nel paragrafo precedente. La classificazione viene riportata nella tabella 1 (chiamata *matrice di classificazione*), in cui sui due assi principali si trovano rispettivamente la presenza o meno delle componenti di trend e stagionalità. Tuttavia, preliminarmente alla classificazione basata sulla stagionalità, vi è la distinzione tra serie sporadiche e serie continue, discussa nel precedente paragrafo.

		Presenza di trend		
		assenza di trend	trend lineare	trend non lineare
sporadica	intermittente	<i>stazionaria</i>	<i>crescente/ decrescente</i>	<i>crescente/ decrescente</i>
	stagionale	<i>stazionaria</i>	<i>crescente/ decrescente</i>	<i>crescente/ decrescente</i>
continua	non stagionale	<i>stazionaria</i>	<i>crescente/ decrescente</i>	<i>crescente/ decrescente</i>
	stagionale	<i>stazionaria</i>	<i>crescente/ decrescente</i>	<i>crescente/ decrescente</i>

Tabella 1.1 Classificazione delle serie storiche di domanda

Per ciascun incrocio della *matrice di classificazione* delle serie storiche si ottiene una tipologia base di serie storica; le serie non stazionarie possono presentare trend crescente o decrescente, approssimabile tramite un'opportuna funzione matematica.

Alcuni esempi di serie storiche possono essere i seguenti grafici:

- serie storica continua non stagionale, caratterizzata da trend lineare crescente;

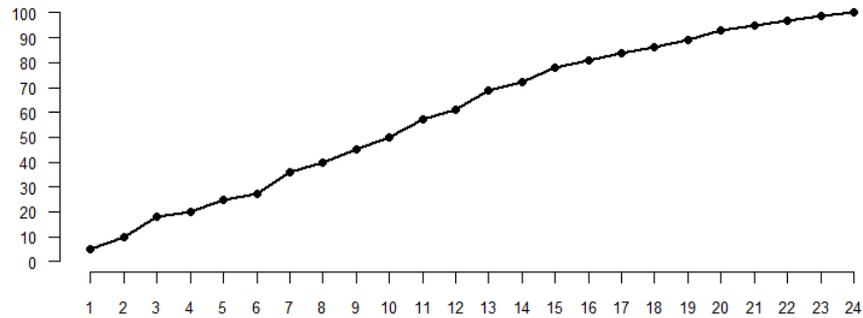


Figura 1.1: Trend lineare

- serie di tipo stagionale, avente lunghezza pari a 6 periodi, caratterizzata da trend lineare crescente ;

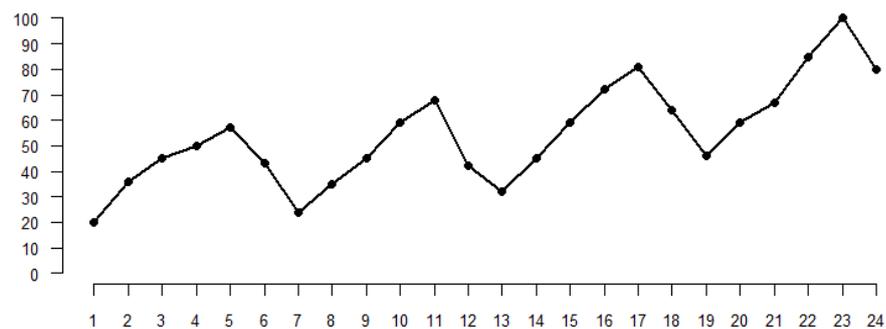


Figura 1.2: Serie storica continua stagionale con trend lineare crescente

- serie sporadica stazionaria, caratteristica di prodotti ordinati poco frequentemente ed in volumi irregolari (ad esempio: beni di lusso, parti di ricambio del settore automobilistico);

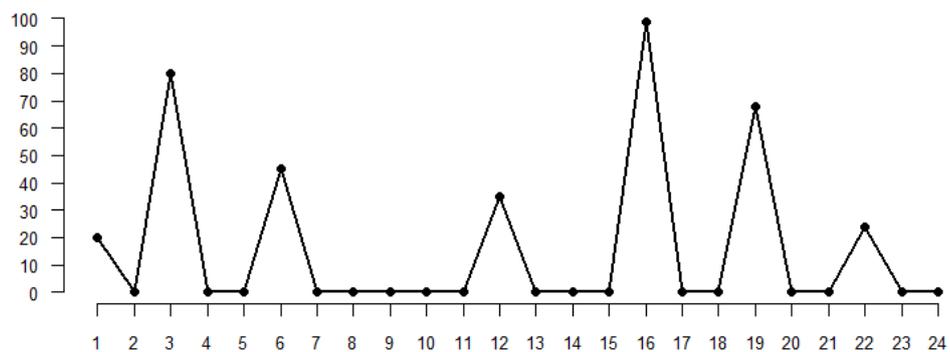


Figura 1.3. Serie storica sporadica non stagionale né affetta da trend

- serie sporadica stagionale stazionaria, caratteristica di prodotti venduti in singole stagioni di vendite periodiche (ad esempio prodotti per l'abbigliamento estivo o prodotti di alta moda)

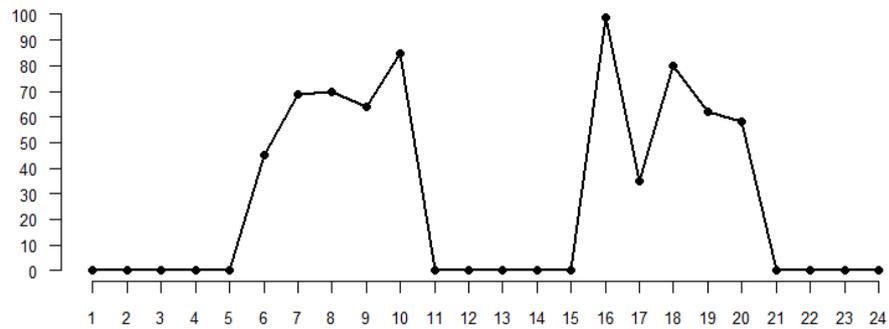


Figura 1.4: Serie storica sporadica stagionale stazionaria

- serie continua non stagionale crescente in modo parabolico

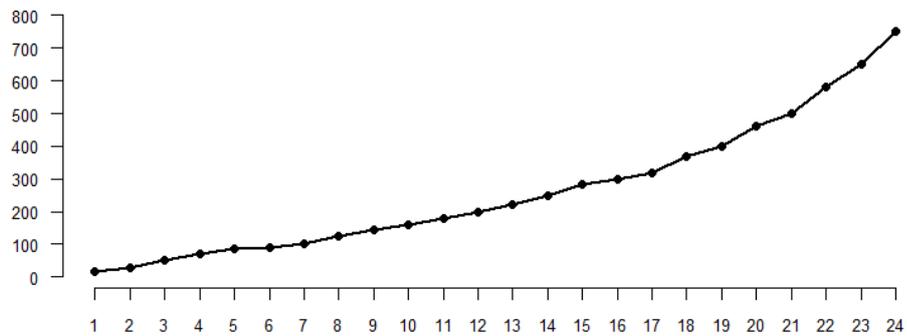


Figura 1.5: Serie storica continua non stagionale parabolicamente crescente

2. Sales Cleaning

Il procedimento di *sales cleaning* è essenziale per ricavare la baseline storica D_t^B da fornire come input ai modelli di sales forecasting; ed in generale è composto dai seguenti passi:

- 1) Individuazione del periodo da analizzare (serie storica ridotta senza picchi promozionali)
- 2) Definizione dei parametri di calcolo delle procedure di sales cleaning
- 3) Calcolo e rimozione delle componenti di stagionalità e trend dalla serie ridotta
- 4) Certificazione e modifica degli outliers della serie storica ridotta tramite algoritmi di sales cleaning
- 5) Aggiunta delle componenti di stagionalità e trend alla serie storica ridotta rettificata
- 6) Modifica dei valori di vendite regolari della serie completa con quelli della serie ridotta rettificata
- 7) Calcolo e rimozione delle componenti di stagionalità e trend dalla serie completa
- 8) Calcolo e rimozione impatto promozionale
- 9) Aggiunta di trend e stagionalità alla serie completa per ottenere la baseline finale

In generale, data una serie storica continua di valori di domanda D_t nell'orizzonte temporale pianificato $t = 1, \dots, T$ si procede alla determinazione delle componenti regolari di trend T_t e stagionalità S_t e successivamente alla pulizia statistica degli outliers o valori anomali.

Questi ultimi in una serie storica possono essere attribuiti a quantità elevate, molto superiori al valor medio della serie, ad esempio vendite occasionali difficilmente riproponibili in futuro o all'opposto a mancate vendite dovute alla rottura di stock (stockout) in cui il prodotto non era disponibile a scaffale presso il negozio. Inoltre gli outliers più comuni sono rappresentati da attività promozionali che spingono la domanda verso l'alto, e verranno studiati più approfonditamente nel seguito.

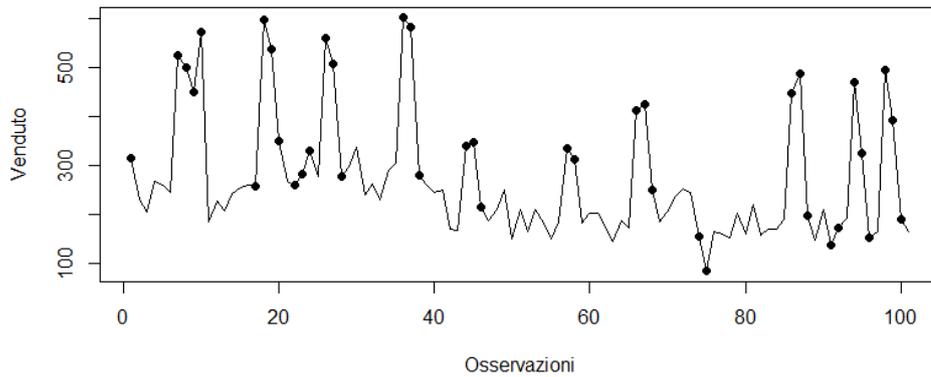


Figura 2.1: Serie storica originale (D_t) con eventi promozionali evidenziati dai pallini

Si procede ad individuare il periodo di analisi della serie storica D_t suddividendola in due componenti: la *domanda regolare* (D_t^R), dove la vendita è indipendente dalla presenza di promozioni, e la *domanda promozionale* (D_t^P) dove invece ne è inficiata. In formule risulta:

$$D_t = \begin{cases} D_t^R & \text{se } t \neq \text{promozione} \\ D_t^P & \text{se } t = \text{promozione} \end{cases} \quad (2.1)$$

Si procede considerando nell'analisi solo D_t^R , ovvero i periodi non soggetti ad attività promozionali della serie storica, ponendoli uno di seguito all'altro in modo da creare una serie storica con una scala temporale fittizia, raffigurata nel grafico 2.2.

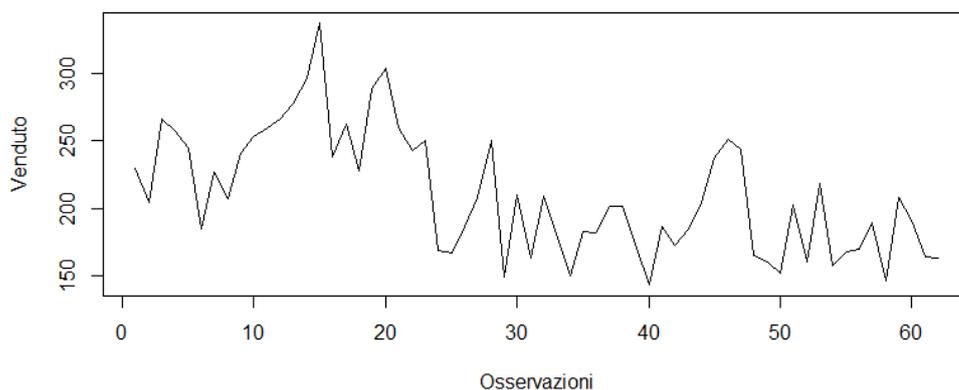


Figura 2.2: Serie regolare (D_t^R)

A questo punto si scompongono e rimuovono dalla serie storica ristretta le componenti riguardanti la stagionalità S_t e il trend T_t attraverso due metodologie principali: *scomposizione moltiplicativa* oppure *scomposizione additiva*.

2.1 Il metodo della scomposizione moltiplicativa

Il modello moltiplicativo per la serie D_t^R in formole può essere espresso come:

$$D_t^R = S_t * T_t * C_t * M_t \quad (2.2)$$

Le fasi dell'algorithmo di scomposizione moltiplicativa per una serie continua e regolare D_t^R , estesa nell'intervallo $t = 1, \dots, T$, sono le seguenti:

- 1) identificazione della componente congiunta di trend e ciclicità: $T_t * C_t$
- 2) identificazione della componente congiunta di stagionalità ed oscillazione casuale intorno alla media: $S_t * M_t$
- 3) isolamento della componente di stagionalità S_t
- 4) destagionalizzazione della serie storica originale D_t^R
- 5) isolamento della componente di tendenza T_t
- 6) isolamento della componente di ciclicità C_t
- 7) isolamento della componente di fluttuazione casuale M_t

Fase 1. Per ottenere la componente congiunta moltiplicativa di trend e ciclicità, pari a $T_t * C_t$ si possono utilizzare due approcci: *media mobile centrata* e metodo delle *differenze prime*.

L'*algorithmo di media mobile centrata* (Centered Moving Average, CMA) è una variante della media mobile semplice, tale per cui i valori inclusi nella media aritmetica sono equamente distribuiti a destra e a sinistra del valore corrente t -esimo della serie storica. In questo modo la media risulta centrata attorno al valore di domanda D_t^R . Per utilizzarlo bisogna porre come valore di ampiezza della media centrale la lunghezza $\tau = L$ del ciclo stagionale di vendita. In formole il calcolo della media mobile centrata di semi ampiezza $\tau/2$ risulta:

$$CMA_t(\tau) = \frac{1}{\tau+1} * \sum_{\vartheta=t-\tau/2}^{t+\tau/2} D_{\vartheta}^R \quad (2.3)$$

ipotizzando che il valore $\tau/2$ sia intero. La media centrata è eseguita su un totale di $\tau+1$ osservazioni: $\tau/2$ valori a sinistra in aggiunta ai $\tau/2$ valori a destra e al valore centrale D_t^R

Nel caso in cui τ risulti un numero pari (ad esempio 12 mesi o 52 settimane per stagionalità annuali), si utilizza una procedura di esecuzione di due medie mobili, ovvero la seguente formula:

$$CMA_t(\tau) = \frac{D_{t+\frac{\tau}{2}} + D_{t+(\frac{\tau}{2})-1} + \dots + D_{t-(\frac{\tau}{2})+1}}{2\tau} + \frac{D_{t+(\frac{\tau}{2})-1} + D_{t+(\frac{\tau}{2})-2} + \dots + D_{t-(\frac{\tau}{2})}}{2\tau} \quad (2.4)$$

La componente congiunta di trend e ciclicità cercata viene quindi determinata, per ciascun periodo della serie D_t^R , tramite la seguente relazione:

$$T_t * C_t = CMA_t(\tau = L) \quad (2.5)$$

Un approccio più semplice e immediato per il calcolo della componente congiunta di trend e ciclicità è il *metodo delle differenze prime*. Dove, definita una differenza del primo ordine pari alla differenza fra la domanda storica del periodo corrente D_t^R e quella del periodo precedente D_{t-1}^R , in formule risulta:

$$T_t * C_t = D_t^R - D_{t-1}^R \quad (2.6)$$

Fase 2. La componente congiunta di stagionalità S_t e rumore casuale attorno al valor medio M_t si ottiene dal rapporto fra la serie di domanda D_t^R e il prodotto fra le componenti moltiplicative di trend e ciclicità:

$$S_t * M_t = \frac{D_t^R}{T_t * C_t} = \frac{D_t^R}{CMA_t(L)} \quad (2.7)$$

dove $T_t * C_t$ è stata calcolata con il metodo della media mobile centrata; oppure:

$$S_t * M_t = \frac{D_t^R}{T_t * C_t} = \frac{D_t^R}{D_t^R - D_{t-1}^R} \quad (2.8)$$

dove $T_t * C_t$ è stata calcolata con il metodo delle differenze prime.

Fase 3. La componente di stagionalità depurata dalla componente casuale di domanda, viene calcolata tramite la media aritmetica dei valori della componente congiunta $S_t * M_t$ per periodi di tempo omologhi rispetto alla lunghezza del ciclo stagionale di vendita.

Ad esempio i mesi di Gennaio vengono mediati fra loro, nel caso in cui le osservazioni analizzate siano mensili ed il ciclo di vendita annuale.

Per ciascun istante del ciclo stagionale $l = 1, \dots, L$ (ciascun mese in un anno, o ciascuna settimana in un trimestre, ecc,) viene isolata la componente stagionale S_l definita come:

$$S_l = \frac{\sum_{t \in K(l)} S_t * M_t}{\text{card} \{K(l)\}} \quad (2.9)$$

dove $K(l)$ è l'insieme che contiene i periodi omologhi di tipo l (per esempio Gennaio 2011, Gennaio 2012, ...).

I coefficienti S_l definiscono la percentuale di vendite al di sopra o al di sotto della media nel corso delle stagioni di vendita analizzate nella serie storica.

Vale la relazione: $S_l = S_{l+L}$ per ogni periodo omologo $l = 1, \dots, L$

Fase 4. Per ottenere la serie storica destagionalizzata si procede eliminando la componente stagionale dalla serie storica di partenza tramite il rapporto fra la serie D_t^R e la componente stagionale isolata nella fase 3. In formule risulta:

$$DS_t = \frac{D_t^R}{S_l(t)} = T_t * C_t * M_t \quad (2.10)$$

essendo $S_l(t)$ il coefficiente stagionale del periodo t , di classe l .

Fase 5. La componente di trend T_t , ipotizzata di tipo lineare, viene ottenuta interpolando i valori della serie storica destagionalizzata DS_t (ricavata nella fase 4 del metodo moltiplicativo) con il tempo $t = 1, \dots, T$, ovvero applicando un modello di regressione lineare semplice, dove:

- la serie storica destagionalizzata DS_t è la variabile dipendente di output
- la sequenza temporale $t = 1, 2, 3, \dots, T$ è la variabile indipendente in input

Si ottiene quindi in funzione del tempo t e dei coefficienti a e b della retta di regressione, stimati mediante il metodo dei minimi quadrati, l'espressione lineare per la componente di tendenza; che in formule risulta:

$$T_t = a * t + b \quad (2.11)$$

Va sottolineato il fatto che tale relazione viene applicata nel passato come approssimazione della componente di trend, il quale si suppone continuare negli istanti futuri, indagati dai modelli di *sales forecasting*.

Fase 6. La componente di ciclicità C_t viene ottenuta dal rapporto dei valori di domanda ottenuti mediante il calcolo della media mobile centrata, relativi alla componente congiunta di trend e ciclicità, rispetto alla retta di regressione della sola componente di tendenza. In formule risulta:

$$C_t = \frac{T_t * C_t}{T_t} = \frac{CMA_t(L)}{a * t + b} \quad (2.12)$$

Fase 7. L'ultima fase dell'algoritmo prevede la determinazione della componente erratica M_t della serie originaria. Dopo aver destagionalizzato e detrendizzato la serie storica D_t^R ottenendo la nuova serie adimensionale:

$$D_t^* = \frac{D_t^R}{S_l(t) * T_t} = \frac{D_t^R}{S_l(t) * (a * t + b)} \quad (2.13)$$

si procede determinando la componente erratica M_t come:

$$M_t = m * D_t^* = \frac{m * D_t^R}{S_l(t) * T_t} \quad (2.14)$$

essendo m la media aritmetica di tutti i valori della serie storica D_t^R , calcolata su tutti i periodi storici $t = 1, \dots, T$.

2.2 Il metodo della scomposizione additiva

In analogia con il metodo della scomposizione moltiplicativa, il modello additivo di scomposizione di una serie storica continua e regolare per la serie D_t^R può essere espresso dalla seguente equazione:

$$D_t^R = S_t + T_t + C_t + M_t \quad (2.15)$$

dove i diversi contributi regolari della domanda sono sommati linearmente. Le fasi dell'algoritmo di scomposizione additiva per una serie storica D_t^R , estesa nell'intervallo

$t = 1, \dots, T$, sono le seguenti:

1. identificazione della componente congiunta di trend e ciclicità: $T_t + C_t$
2. identificazione della componente congiunta di stagionalità ed oscillazione casuale attorno alla media: $S_t + M_t$
3. isolamento della componente di stagionalità S_t
4. isolamento della componente di erraticità M_t

Per il calcolo della componente congiunta additiva di trend e ciclicità può essere utilizzato il metodo della media mobile centrata $CMA_t(L)$ dove L è pari alla lunghezza del ciclo stagionale.

Nota la componente $T_t + C_t$, la componente congiunta additiva di stagionalità e rumore casuale viene determinata per differenza, nel seguente modo:

$$S_t + M_t = D_t^R - (T_t + C_t) \quad (2.16)$$

Le componenti periodiche di stagionalità, in numero pari a L , si determinano dalla seguente formula:

$$S_l = \frac{\sum_{t \in K(l)} (S_t + M_t)}{\text{card} \{K(l)\}} \quad (2.17)$$

dove $K(l)$ è l'insieme che contiene i periodi omologhi di tipo l (ad esempio Gennaio 2011, Gennaio 2012, ...).

Infine la componente di fluttuazione casuale M_t viene nuovamente isolata per differenza di componenti regolari note. In formule risulta:

$$M_t = D_t^R - (T_t + C_t + S_t) \quad (2.18)$$

In sintesi la differenza fondamentale dei due approcci è il fatto che nella scomposizione moltiplicativa della serie storica di domanda D_t^R , questa viene divisa per le componenti di trend T_t e stagionalità S_t ottenendo:

$$D_t^* = \frac{D_t^R}{T_t * S_t(t)} \quad (2.19)$$

mentre nella scomposizione additiva della serie storica di domanda vengono sottratte le componenti di trend T_t e stagionalità S_t nel seguente modo:

$$D_t^* = D_t^R - (T_t + S_t(t)) \quad (2.20)$$

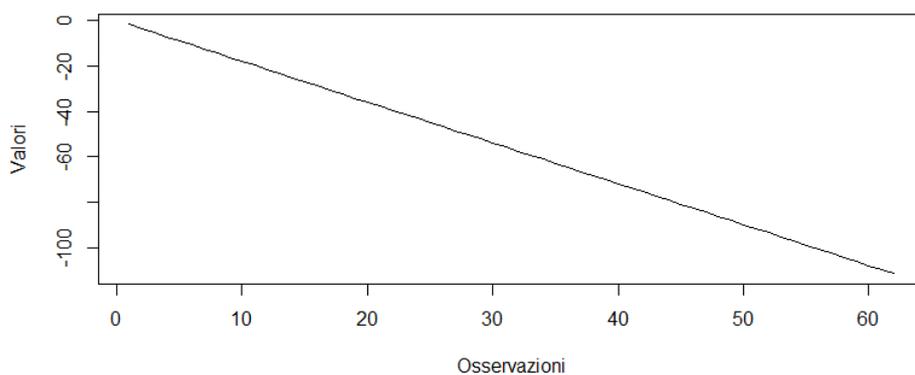


Figura 2.3: Componente di trend lineare decrescente per la serie (D_t^R)

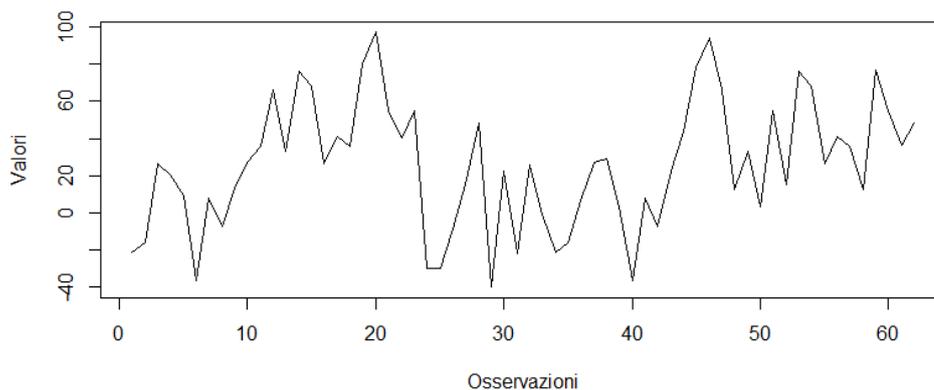


Figura 2.4: Componente stagionale della serie (D_t^R)

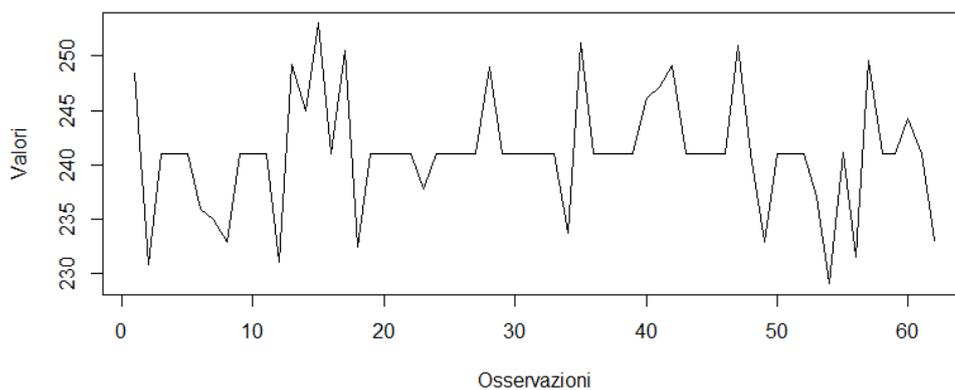


Figura 2.5: Serie regolare detrendizzata e destagionalizzata (D_t^*)

Sia se viene utilizzato il metodo della scomposizione moltiplicativa, sia quello della scomposizione additiva, se la serie storica risultante D_t^* presenta ancora valori anomali è necessario utilizzare degli algoritmi di sales cleaning per modificare tali valori come l'*intervallo di confidenza* o la media mobile, in modo da ottenere la serie storica D_t^C con i valori rettificati.

2.3 Metodo dell'intervallo di confidenza per la rettifica degli outliers

Il metodo dell' *intervallo di confidenza (IC)* sfrutta il calcolo preliminare della media μ_D e la deviazione standard σ_D della serie di domanda detrendizzata e destagionalizzata D_t^* dove $t=1,..,T$, che vengono utilizzate nella costruzione dell'IC all' α % che risulta essere il seguente:

$$[INF_D ; SUP_D] = [\mu_D - k_\alpha * \sigma_D ; \mu_D + k_\alpha * \sigma_D] \quad (2.21)$$

dove k_α è il percentile di ordine α della distribuzione normale standard $N(0,1)$. Quindi, per esempio, per un IC al $\alpha=95\%$ il quantile $k_\alpha= 1.96$.

L'intervallo così costruito gode di buone proprietà, infatti è simmetrico intorno al valore atteso ed ha ampiezza pari a $2 * k_\alpha * \sigma_D$.

Si verifica quindi che tutti i valori storici, presi singolarmente dal meno al più recente, siano contenuti all'interno dell'intervallo individuato. Se così non fosse si procede a modificarlo ponendolo uguale all' estremo inferiore dell'IC (INF_D) (o all'estremo superiore chiamato SUP_D), qualora il valore di domanda fosse rispettivamente più grande (o più piccolo) rispetto ai valori estremi dell'IC.

Il metodo dell'IC , tuttavia, presenta due varianti: *modello statico* o *modello dinamico*.

Il primo consiste nel fatto di calcolare i valori di μ_D e σ_D , rispettivamente media e deviazione standard della serie storica, utilizzando tutti i T periodi di cui la serie è composta.

Tuttavia, così facendo, il metodo risulta debole poiché i parametri sono calcolati senza tenere conto della numerosità e intensità degli outliers, che possono distorcere il calcolo della media e deviazione standard della serie.

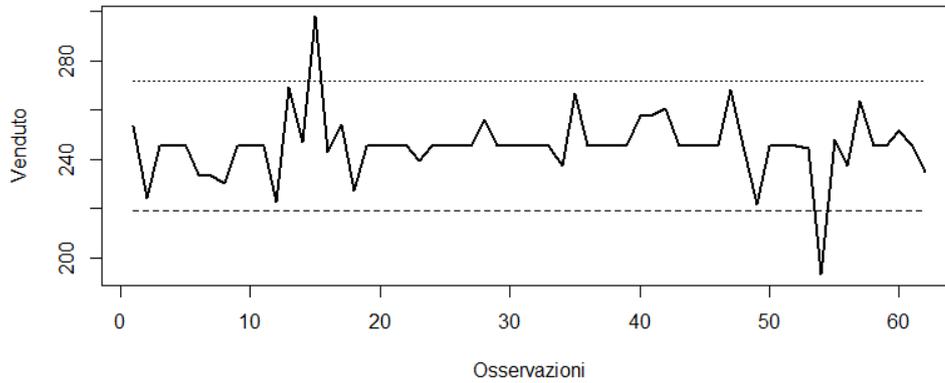


Figura 2.6: Serie D_t^* con IC statico

Il secondo metodo, ovvero il modello dinamico, consiste nel valutare i valori di media e deviazione standard in modo progressivo, col passare del tempo. Infatti i parametri μ_D e σ_D vengono ricalcolati ad ogni $t=1, \dots, T$ considerando solo i primi $t-1$ periodi storici, eventualmente già rettificati. La qualità dei due parametri di input dell'algoritmo migliora progressivamente tenendo conto delle modifiche degli outliers della serie storica.

Ciò comporta il verificarsi di un processo di autoapprendimento poiché il modello migliora la propria capacità di identificare gli outliers e quindi di correggerli.

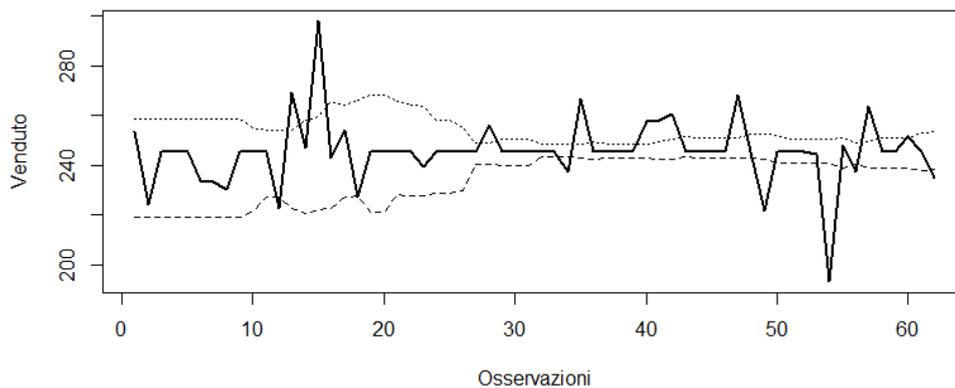


Figura 2.7: Serie D_t^* con IC dinamico

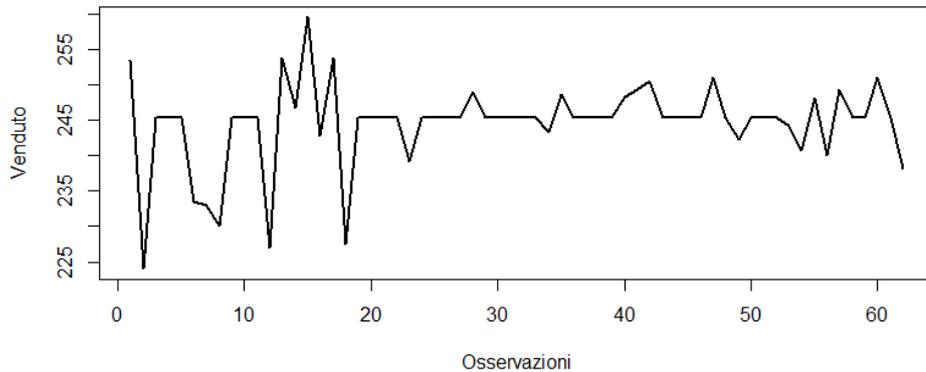


Figura 2.8: Serie corretta dagli outliers con metodo dell' IC dinamico (D_t^{CR})

2.4 Metodo della media mobile per la rettifica degli outliers

In alternativa al metodo dell'intervallo di confidenza si può utilizzare l'algoritmo della *Media Mobile*, in inglese moving average (MA). Questo è caratterizzato dal fatto di eseguire una media aritmetica dei valori di domanda degli ultimi τ periodi rispetto al periodo corrente t , in modo da calcolare l'entità del cleaning per il valore corrente D_t^* .

Anche in questo caso, un valore di domanda D_t^* risulta essere un outlier se cade al di fuori della banda di confidenza, calcolato come funzione di media e deviazione standard della distribuzione di domanda.

Una volta individuati, gli outliers vengono rettificati tramite la definizione della media mobile a τ periodi, per la serie storica di domanda D_t^* , che in formule risulta:

$$MA_t(\tau) = \frac{1}{\tau} * \sum_{\vartheta=t-\tau}^{t-1} D_{\vartheta}^* \quad (2.22)$$

In altre parole gli ultimi τ valori della serie di domanda D_t^* , escluso il valore al tempo t , vengono mediati in modo uniforme, attraverso la media aritmetica dei rispettivi valori.

La serie storica MA_t risultante è caratterizzata da una nuova serie di valori D_t^{CR} , che presentano differenze dalla serie originaria, poiché gli outliers sono stati corretti.

In sintesi l'algoritmo della media mobile, una volta individuato un valore D_t^* come outlier applica il valore MA_t calcolato sulla base degli ultimi τ valori di domanda, senza tenere conto del fatto che gli ultimi τ valori siano stati precedentemente modificati dal medesimo algoritmo.

Poiché la serie storica D_t^* è stata rettificata in D_t^{CR} , si può procedere all'applicazione delle componenti di trend T_t e stagionalità S_t calcolate

precedentemente, per ottenere la baseline di domanda della serie ridotta che verrà poi utilizzata negli algoritmi di *sales forecasting*.

In particolare se è stato usato il metodo della scomposizione moltiplicativa, il valore finale della domanda base regolare (*regular baseline*) è:

$$D_t^{BR} = D_t^{CR} * T_t * S_l(t) \quad (2.23)$$

Mentre se è stato utilizzato il metodo della scomposizione additiva il valore finale di domanda base regolare risulta essere:

$$D_t^{BR} = D_t^{CR} + T_t + S_l(t) \quad (2.24)$$

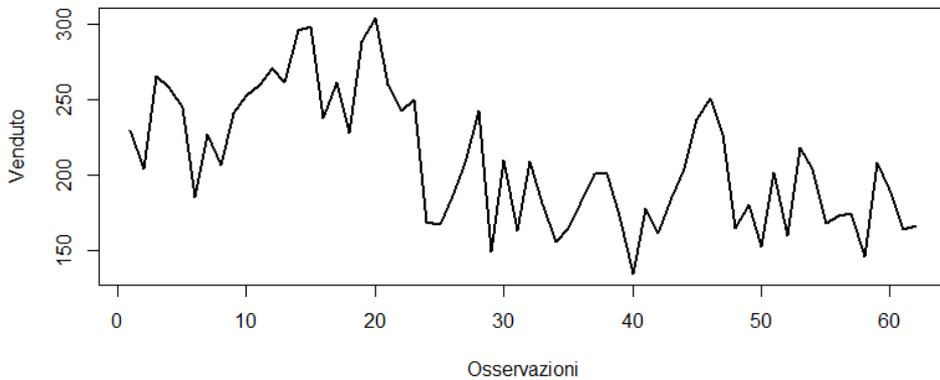


Figura 2.9: Serie D_t^{CR} con aggiunta di stagionalità e trend (D_t^{BR})

Va sottolineato il fatto che se un valore della serie di domanda ridotta D_t^R non risulta classificato come outlier, presenterà lo stesso valore sia prima che dopo l'utilizzo del processo di sales cleaning.

La serie storica di domanda dopo il processo di sales cleaning (solo per la parte di vendite regolari) risulta essere:

$$D_t^{PARZ} = \begin{cases} D_t^{BR} & \text{se } t \neq \text{promozione} \\ D_t^p & \text{se } t = \text{promozione} \end{cases} \quad (2.25)$$

dove D_t^p è il valore della serie storica di domanda al tempo t soggetta esclusivamente ad una promozione di vendita.

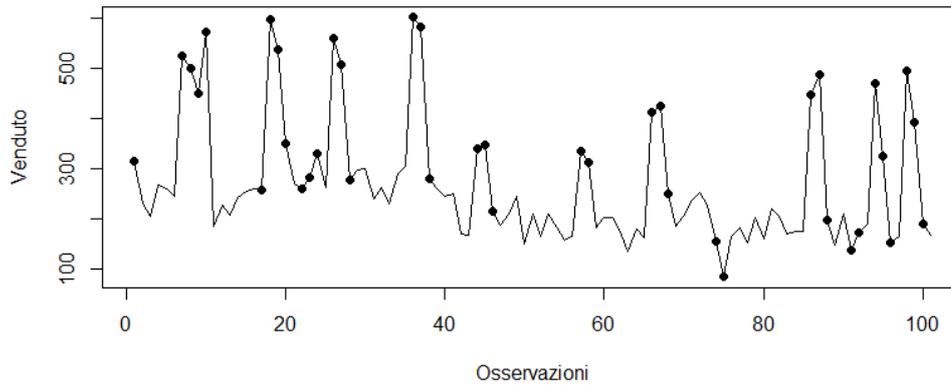


Figura 2.10: Serie D_t^{PARZ} con solo la parte di vendita regolare corretta e con eventi promozionali evidenziati dai pallini

A questo punto la serie D_t^{PARZ} presenta valori rettificati solo per la parte di vendite regolari e quindi non può essere ancora utilizzata dai modelli di sales forecasting. Si procede quindi a isolare gli effetti dei picchi promozionali dalla serie. Per fare ciò si calcola la componente di trend T_t e la componente stagionalità $S_l(t)$ considerando l'intera serie D_t^{PARZ} e con i metodi della scomposizione additiva o moltiplicativa, già visti in precedenza, e si rimuovono quindi tali componenti dalla serie, ottenendo la nuova serie D_t^* .

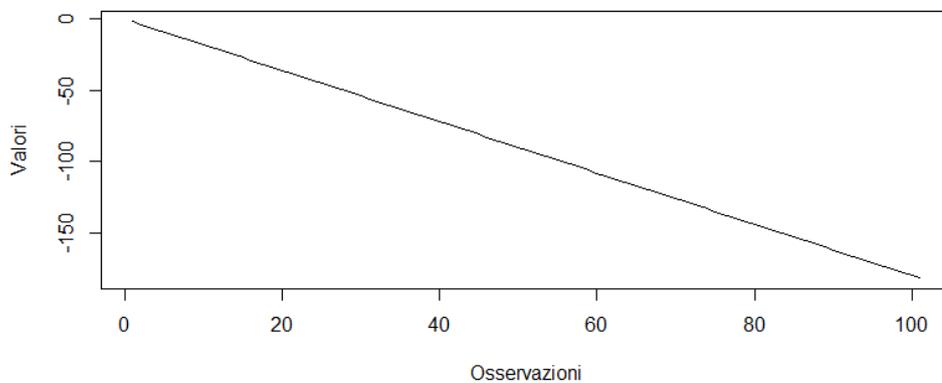


Figura 2.11: Componente di trend lineare decrescente per la serie D_t^{PARZ}

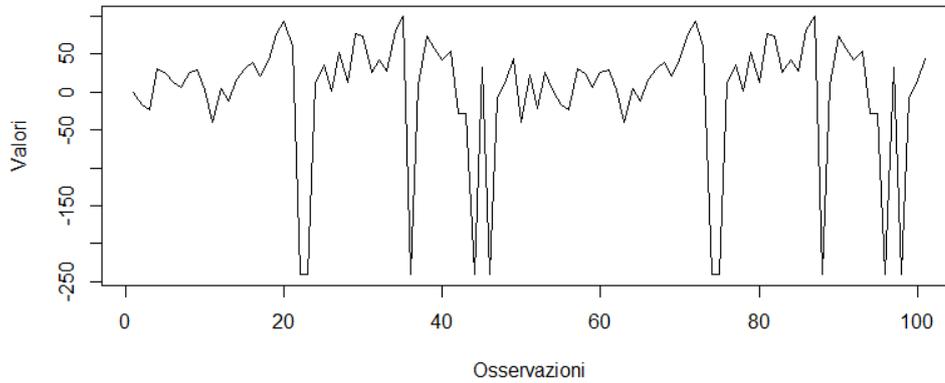


Figura 2.12: Componente di stagionalità per la serie D_t^{PARZ}

Quest'ultima, detrendizzata e destagionalizzata, presenta degli outliers solo in presenza di picchi promozionali.

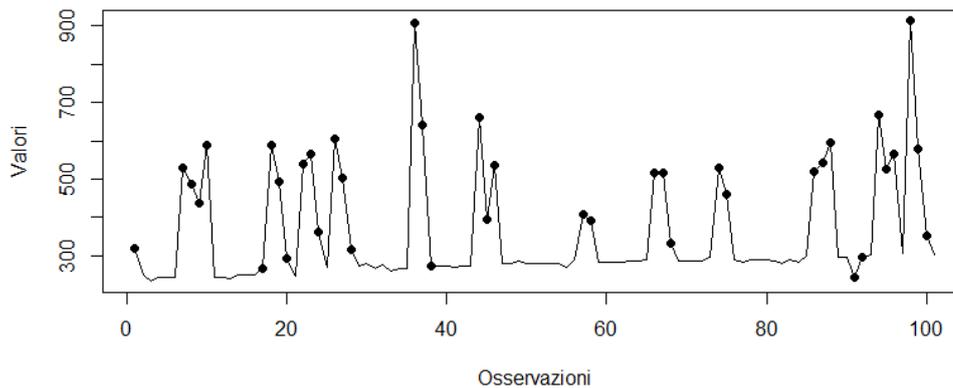


Figura 2.13: Serie D_t^{PARZ} detrendizzata e destagionalizzata (D_t^*)

2.5 Metodo dell'impatto promozionale

Si procede quindi a stimare ed eliminare l'impatto derivante dai vari tipi di promozione Δ_t in modo da avere una serie storica non influenzata dagli outliers promozionali D_t^C . In formule risulta:

$$D_t^C = \frac{D_t^*}{\widehat{\Delta}_t} \quad (2.26)$$

dove $\widehat{\Delta}_t$ è l'impatto promozionale della promozione che deve essere stimato perciò esso viene utilizzato oltre che nel processo di sales cleaning anche in quello di sales forecasting per prevedere l'impatto di una promozione futura².

² Il metodo utilizzato per la stima e la previsione dell'impatto promozionale Δ_t verrà approfondito nel caso di studio nel capitolo 5

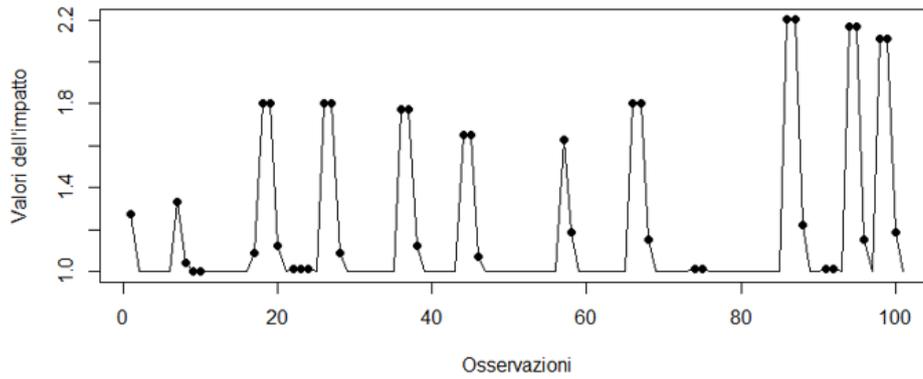


Figura 2.11: Stima dell'impatto promozionale della serie D_t^* con venti promozionali evidenziati dai pallini

Va evidenziato il fatto che nei periodi dove non è presente la promozione l'impatto è $\Delta_t = 1$ e quindi questo implica che:

$$D_t^C \equiv D_t^* \quad \text{se } t \neq \text{promozione} \quad (2.27)$$

Successivamente alla serie D_t^C vengono aggiunte le componenti di trend e stagionalità calcolate precedentemente con uno dei due metodi di scomposizione. Se si è utilizzato il metodo della scomposizione moltiplicativa, si ottiene:

$$D_t^B = D_t^C * T_t * S_l(t) \quad (2.28)$$

mentre se è stato usato il metodo della scomposizione additiva, quello che risulta in formule è:

$$D_t^B = D_t^C + T_t + S_l(t) \quad (2.29)$$

Alla fine del processo di *sales cleaning* la serie storica originale D_t , formata sia dai periodi promozionali (D_t^P) che regolari (D_t^R), viene trasformata nella serie base di domanda D_t^B che può quindi essere utilizzata come input dai modelli di sales forecasting descritti nel capitolo 4.

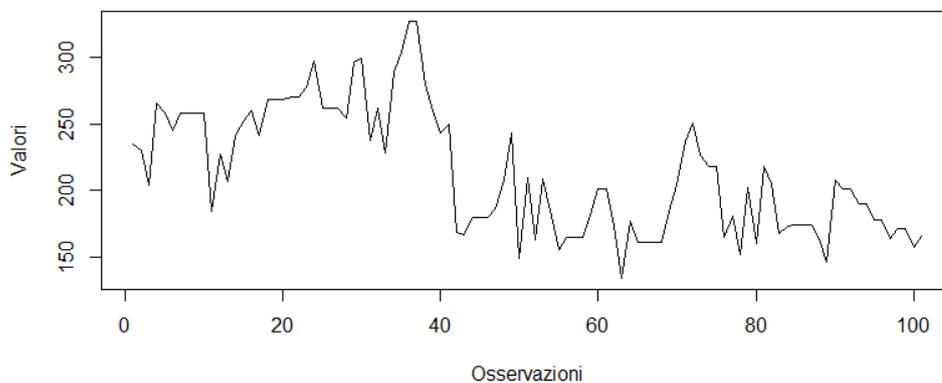


Figura 2.12: Serie di domanda base (D_t^B)

In sintesi il processo di sales cleaning può essere riassunto nel seguente schema logico:

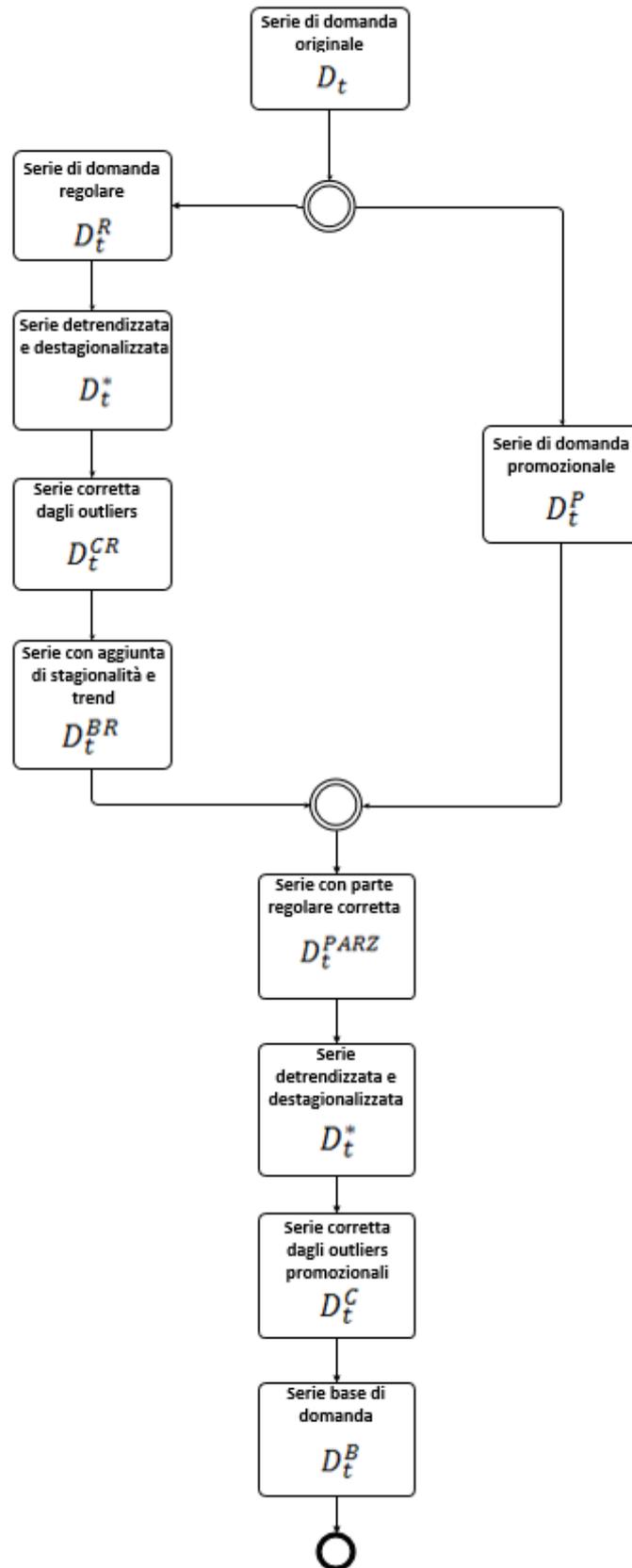


Figura 2.13 Schema logico del processo di sales cleaning

3. Valutazione dei modelli di Sales Forecasting

La serie storica di domanda baseline D_t^B , calcolata grazie al processo di sales cleaning, viene utilizzata dai modelli di *sales forecasting* per generare la previsione della domanda commerciale per gli SKU nella finestra temporale pianificata.

Questi modelli estrapolano i valori futuri di domanda previsionale come proiezione delle regolarità identificate nel passato. In particolare attribuiranno molta importanza ai picchi promozionali avvenuti nel passato oltre alle componenti regolari della domanda, se la previsione riguarda periodi promozionali, mentre se è una previsione della domanda base (baseline) senza effetto di promozioni, presteranno solo attenzione a componenti di stagionalità e trend.

Va evidenziato il fatto che le previsioni \hat{y}_t prodotte dai modelli statistici riguardano esclusivamente periodi di vendita regolari (poiché la baseline D_t^B presenta solo vendite regolari), mentre se si è interessati a previsioni per periodi promozionali si deve aggiungere alla previsione \hat{y}_t l'impatto promozionale stimato $\widehat{\Delta}_t$ attraverso il metodo che si è utilizzato per il calcolo di quest'ultimo nel processo di sales cleaning. In altre parole si utilizza la seguente formula:

$$\hat{y}_t = \begin{cases} \hat{y}_t & \text{se } t \neq \text{promozione} \\ \hat{y}_t * \widehat{\Delta}_t & \text{se } t = \text{promozione} \end{cases} \quad (3.1)$$

In letteratura esistono molti modelli per il sales forecasting, tutti accomunati dal fatto di avere un'*accuratezza previsionale (forecast accuracy)* che può essere utilizzata per confronti tra i vari modelli. Infatti l'accuratezza misura la differenza tra i valori di domanda previsti e quelli a consuntivo, ovvero che si sono realizzati effettivamente e sono disponibili nel DWH.

Si definisce *forecast error* la quantità:

$$e_t = y_t - \hat{y}_t \quad \text{per } t = 1, \dots, T \quad (3.2)$$

dove $y_t \equiv D_t$ è la domanda effettiva al tempo t , mentre \hat{y}_t è la previsione effettuata al tempo t .

L'errore di previsione, per costruzione, sarà positivo nei casi di sottostima della vendite reali; mentre assumerà un valore negativo dove il forecast è più elevato della domanda effettiva (sovrastima).

Nei casi in cui l'errore di previsione è nullo significa che la previsione coincide perfettamente con la domanda effettiva registrata.

Poiché è molto difficile comprendere se un metodo previsionale sia adatto a fornire una buona stima dei dati analizzati, è auspicabile compiere dei test sulla

forecast accuracy dei risultati ottenuti da tali metodi. L'accuratezza delle previsioni può essere valutata solo considerando quanto bene ogni modello si adatta su dati che non sono stati usati per generare la stima di quello specifico modello e non quanto il modello sembri adattarsi ai dati storici osservati in precedenza (Hyndman, 2014).

Per questo motivo, quando si sceglie quale metodo previsionale utilizzare, è consuetudine dividere i dati a disposizione, ovvero la serie storica di domanda. Infatti a seconda dell'obiettivo predittivo dei Demand Planner (previsioni di domanda regolare o promozionale) si considerano due metodi per suddividere la serie storica di domanda con cui calcolare la previsione e la relativa accuratezza previsionale.

3.1 Training set e test set standard

Se l'obiettivo è quello di prevedere la domanda commerciale regolare, per valutare l'accuratezza delle previsioni si preferisce utilizzare uno specifico intervallo temporale di verifica (*test set*), che non è altro che un sottoinsieme della serie storica di domanda, tra l'intervallo di stima (*training set*) e quello di *previsione*.

Poiché l'orizzonte temporale della serie di domanda si estende da 1 a T , l'intervallo di training è pari a T_A mentre l'estensione del test set è pari a $T - T_A$. L'intervallo di previsione cade cronologicamente dopo dell'ultimo istante temporale che presenta un valore della serie di domanda, quindi si vuole prevedere la domanda negli istanti futuri ($T + 1, T + 2, T + 3, \dots$)

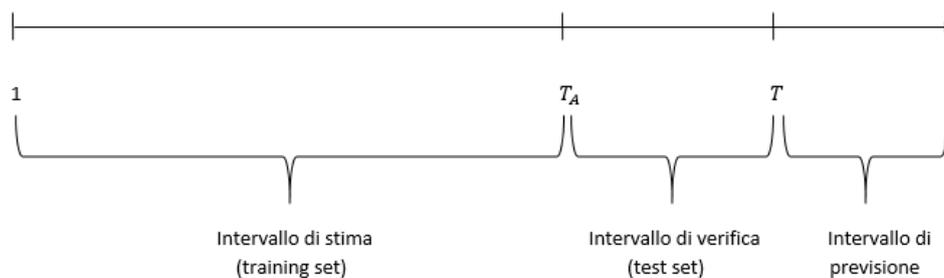


Figura 3.1: Schema di suddivisione della serie storica di domanda in intervallo di stima, di verifica e di previsione.

Il training set, primo dei tre intervalli, viene utilizzato dai modelli per l'apprendimento delle caratteristiche della serie di domanda. Maggiore è l'intervallo considerato, ovvero più dati storici si hanno a disposizione, più cresce il livello di apprendimento degli algoritmi, in particolare nel calcolo delle componenti di trend e stagionalità.

D'altro canto se la serie risulta troppo corta, ad esempio è inferiore a un anno, una stagionalità per previsione settimanale non può essere calcolata.

Mentre il test set, come già specificato, viene utilizzato per la valutazione dell'accuratezza delle previsioni dei modelli di sales forecasting. In particolare questi ultimi elaboreranno la previsione per gli istanti temporali: $T_A + 1, T_A + 2, \dots, T$. Si determinano quindi misure sintetiche di accuratezza previsionale, per i vari modelli di sales forecasting, costruite confrontando i valori puntuali degli errori calcolati nell'intervallo di test. Nell'ultimo intervallo, in senso cronologico, detto intervallo di previsione, gli algoritmi generano il forecast della futura domanda commerciale. La lunghezza di tale intervallo è variabile e dipende dal contesto aziendale. Nel mondo della GDO, per esempio, i prodotti alimentari e freschi hanno una previsione basata su un orizzonte di pochi giorni, mentre sui prodotti a lunga conservazione anche di alcune settimane. Una valutazione dell'accuratezza del metodo previsionale utilizzando questa prassi è detta "out-of-sample performance".

3.2 One-step-ahead forecasts

In alternativa si può usare il metodo chiamato "one-step-ahead forecasts", che consiste nel calcolare una previsione per il primo periodo successivo all'ultima osservazione presente nel training set; allo step successivo al precedente dataset di stima viene aggiunta un'osservazione e viene calcolata nuovamente una previsione per l'osservazione successiva al nuovo training set; si itera quindi la procedura aggiungendo di volta in volta un'osservazione al dataset di stima precedente producendo un forecast per osservazioni diverse. Alla fine della procedura il test set conterrà un numero di osservazioni pari al numero di previsioni prodotte.

Una valutazione dell'accuratezza del metodo previsionale utilizzando questa prassi è detta "in-sample performance"

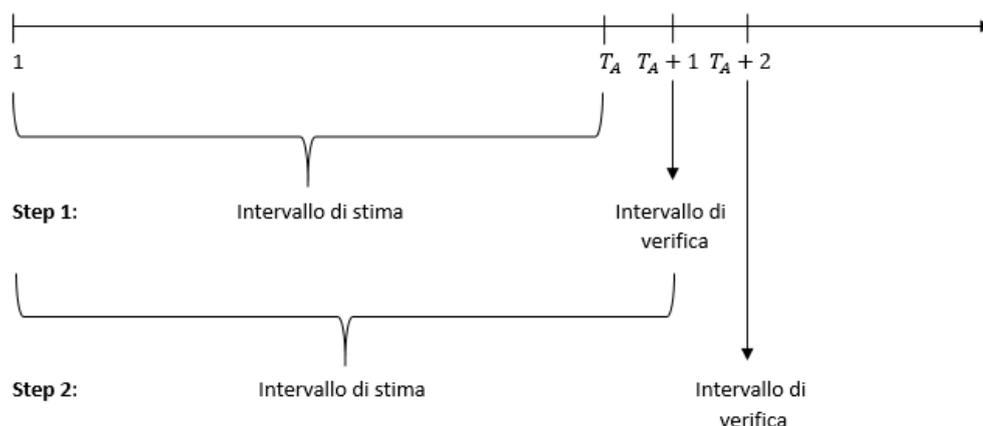


Figura 3.2: One step ahead-forecasts per i primi 2 step

3.3 One-step-promotional-ahead forecasts

Se l'obiettivo è quello di prevedere la domanda commerciale promozionale, l'intervallo di previsione coincide esattamente con quello descritto nella previsione della domanda commerciale regolare, mentre gli intervalli di stima e verifica si moltiplicano. Infatti dato un insieme di picchi promozionali della serie storica (m_1, m_2, \dots, m_M) il metodo chiamato "one-step-promotional-ahead forecasts" (variante del metodo "one-step-ahead forecasts" descritto in precedenza), presenta i seguenti passi principali:

1. m_1 : primo picco promozionale
training set $[1: (m_1-1)]$: ovvero dalla prima osservazione della serie storica a quella precedente il primo picco promozionale (m_1)
test set $[m_1]$: primo picco promozionale in ordine cronologico
2. m_2 : secondo picco promozionale
training set $[(m_1 + 1) : (m_2 - 1)]$: dalla prima osservazione della serie storica dopo il primo picco a quella precedente il secondo picco promozionale
test set $[m_2]$: secondo picco promozionale in ordine cronologico
3.
4. m_M : ultimo picco promozionale
training set $[(m_{M-1} + 1) : (m_M - 1)]$: dalla osservazione successiva al penultimo picco a quella precedente l'ultimo picco promozionale
test set $[m_M]$: ultimo picco promozionale in ordine cronologico

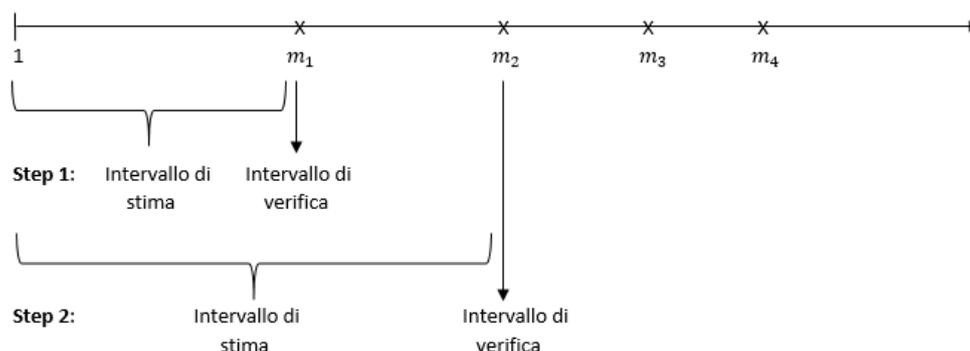


Figura 3.3: One step ahead-promotional-forecasts per i primi 2 step

Alla fine della procedura si avranno quindi M intervalli di stima e M di verifica. Poiché questi ultimi vengono utilizzati con il fine di calcolare una misura di accuratezza previsionale, si avranno anche M errori di previsione puntuale (uno per ciascun intervallo di test e per ciascun modello). Si cercherà quindi di sintetizzare questi M errori utilizzando una delle misure globali di *forecast accuracy* che verranno presentate nel prossimo capitolo, per misurare l'accuratezza previsionale del modello di sales forecasting utilizzato.

3.4 Misure di accuratezza previsionale

In generale si calcola l'errore puntuale di previsione negli istanti temporali all'interno del test set³, utilizzando $f(e_t)$ ovvero un funzione della (3.2). Successivamente l'errore puntuale di previsione viene consolidato attraverso l'utilizzo di una funzione dell'errore totale di stima $f(e_{TOT})$.

Si definisce, percentualmente, forecast accuracy (FA) come il complemento a 1 dell'errore totale di previsione, in formula: $FA = 1 - f(e_{TOT})$

Quindi, per esempio, commettere un errore percentuale di previsione della domanda commerciale del 10%, nel test set, significa produrre un forecast con un'accuratezza pari al 90%.

Il forecast error può essere calcolato *puntualmente*, ovvero sui singoli istanti temporali del test set, oppure *globalmente* sull'intero intervallo di test. Se la differenza riguarda la previsione e il venduto in termini reali e quantitativi, si parla di forecast error *assoluto*; se tale differenza è in termini percentuali si parlerà di forecast error *relativo*.

Le misure globali di errore che pesano l'errore di previsione con il proprio segno, vengono dette *misure di distorsione* e sono molto utilizzate per tener conto dei casi di sovrastima e di sottostima.

In alternativa si possono utilizzare *misure di dispersione* per valutare l'errore di previsione in valore assoluto in modo da evitare la compensazione tra errori di segno opposto.

Le principali misure di forecast error assoluto calcolato puntualmente all'interno del test set sono:

- **Absolute Error:** $AE_t = |e_t| = |y_t - \hat{y}_t|$ (3.3)

- **Squared Error:** $SE_t = e_t^2 = (y_t - \hat{y}_t)^2$ (3.4)

Una misura di forecast error percentuale calcolato puntualmente all'interno del test set è:

- **Percentage Error:** $PE_t = 100 * \frac{e_t}{y_t} = 100 * \frac{y_t - \hat{y}_t}{y_t}$ (3.5)

Questa misura ha il vantaggio di essere indipendente dalla scala della serie storica trattata, mentre ha lo svantaggio di divergere a infinito, fine a essere indefinita per $y_t \rightarrow 0$

³ Il test set indicato in questo capitolo fa riferimento sia all'unico test set utilizzato se l'obiettivo è quello di prevedere la domanda commerciale regolare, sia all'insieme degli M test set del metodo del rolling per picchi promozionali se l'obiettivo è la stima della domanda commerciale promozionale

Tra le *misure globali di distorsione* presenti in letteratura si elencano:

- **Mean Error:**

$$ME = \frac{1}{T-T_A} * \sum_{t=T_A+1}^T e_t = \frac{1}{T-T_A} * \sum_{t=T_A+1}^T (y_t - \hat{y}_t) \quad (3.6)$$

indica se le previsioni sono distorte per eccesso o difetto, dovrebbe assumere un valore vicino a zero

- **Mean Percentage Error:**

$$MPE = \frac{100}{T-T_A} * \sum_{t=T_A+1}^T PE_t = \frac{100}{T-T_A} * \sum_{t=T_A+1}^T \frac{(y_t - \hat{y}_t)}{y_t} \quad (3.7)$$

indica se le previsioni sono distorte per eccesso o difetto ma in termini percentuali, dovrebbe assumere un valore vicino a zero

Tra le *misure di dispersione* presenti in letteratura si individuano:

- **Mean Absolute Error:**

$$MAE = \frac{1}{T-T_A} * \sum_{t=T_A+1}^T AE_t = \frac{1}{T-T_A} * \sum_{t=T_A+1}^T |y_t - \hat{y}_t| \quad (3.8)$$

è la media dei valori assoluti degli errori, più tollerante rispetto a grandi errori occasionali poiché gli errori non vengono elevati al quadrato

- **Mean Absolute Percentage Error:**

$$MAPE = \frac{100}{T-T_A} * \sum_{t=T_A+1}^T |PE_t| = \frac{100}{T-T_A} * \sum_{t=T_A+1}^T \frac{|y_t - \hat{y}_t|}{y_t} \quad (3.9)$$

è più appropriato se i dati presentano trend o stagionalità

- **Root Mean Squared Error:**

$$RMSE = \sqrt{\frac{1}{T-T_A} * \sum_{t=T_A+1}^T (y_t - \hat{y}_t)^2} \quad (3.10)$$

penalizza maggiormente grandi errori rispetto ai piccoli perché prima li eleva al quadrato; è approssimabile alla deviazione standard degli errori se la media degli errori è vicino a zero

Considerando le seguenti relazioni:

- ❖ **Sum of Squared Error:**

$$SSE = \sum_{t=T_A+1}^T SE_t = \sum_{t=T_A+1}^T (y_t - \hat{y}_t)^2 \quad (3.11)$$

- ❖ **Mean Squared Error:**

$$MSE = \frac{SSE}{1-T_A} = \frac{1}{T-T_A} * \sum_{t=T_A+1}^T (y_t - \hat{y}_t)^2 \quad (3.12)$$

Vale la relazione:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{SSE}{1-T_A}} \quad (3.13)$$

Nella maggior parte dei casi vengono preferite le misure di distorsione rispetto alle due misure globali di distorsione presentate, ovvero il ME e il MPE. Quest'ultime infatti non sono molto importanti poiché la distorsione risulta essere piccola quando i parametri dei modelli sono stimati minimizzando l'errore quadratico medio (MSE).

Il MAE e il MAPE hanno il vantaggio di essere compresi più facilmente anche da non specialisti. Queste misure sono meno sensibili all'effetto di grandi outliers e così potrebbero fornire una migliore stima della grandezza di un errore medio quando la distribuzione degli errori non è approssimabile da una normale (Nau, 2014).

In particolare il MAPE è una misura di accuratezza che non è legata alla scala della serie storica, per cui può essere utilizzato per comparare metodi previsionali basati su serie storiche con scala diversa.

La distorsione del MAPE è dovuta al fatto che tale misura non è simmetrica, ovvero scambiare y_t con \hat{y}_t non porta allo stesso risultato, nonostante il fatto che l'errore assoluto AE_t non cambia dopo questo scambio; l'asimmetria è dovuta al denominatore della formula dove scambiare y_t e \hat{y}_t porta a un differente risultato (Tofallis, 2015).

Proprio a causa di questa asimmetria il MAPE penalizza maggiormente le sovrastime ($\hat{y}_t > y_t$) rispetto alle sottostime ($\hat{y}_t < y_t$). Ciò è evidenziato dal range di valori che il MAPE può assumere, ovvero i valori compresi tra $(-100\%, +\infty)$ (Armstrong e Collopy, 1992).

Infine l'RMSE, è una misura di accuratezza che si basa sulla scala dei dati analizzati, ed è quindi poco indicata per confrontare serie storiche su scale diverse. L' RMSE ha il vantaggio di essere una misura ideale per prendere decisioni aziendali, infatti è semplice interpretarlo: un valore inferiore dell' RMSE solitamente implica un risultato migliore. Inoltre, l' RMSE è una buona misura per confrontare modelli in cui la media dell'errore non sia nulla, perché penalizza le distorsioni derivanti da quest'ultima (Nau, 2014).

Va notato come l' RMSE fosse la misura di accuratezza storicamente maggiormente utilizzata prima dell'introduzione del MAPE, più per il fatto di essere usata per la sua popolarità tra gli accademici, piuttosto che per la sua efficienza nello scegliere il metodo di previsione migliore; infatti proprio per questo viene aspramente criticata da Ahlburg (Armstrong, 2008).

Le misure di valutazione dell'errore di previsione possono essere utilizzate per:

- scegliere la migliore combinazione di parametri numerici di ciascun modello di sales forecasting tali per cui ne massimizzano la forecast accuracy, testando la qualità della previsione sull'intervallo di test
- scegliere il migliore modello, ovvero quello che minimizza una prestabilita metrica di errore (o combinazione di esse) nel test set, in modo da ottenere la migliore previsione di domanda per ciascuno SKU

In generale per ciascun SKU è disponibile quindi un insieme di misure di errore riassumibili in un cruscotto di misure di errore come il seguente:

	ME	MPE	MAE	MAPE	RMSE
Target error	ME*	MPE*	MAE*	MAPE*	RMSE*
<i>SKU₁</i>	<i>ME₁</i>	<i>MPE₁</i>	<i>MAE₁</i>	<i>MAPE₁</i>	<i>RMSE₁</i>
<i>SKU₂</i>	<i>ME₂</i>	<i>MPE₂</i>	<i>MAE₁</i>	<i>MAPE₂</i>	<i>RMSE₂</i>
...
<i>SKU_n</i>	<i>ME_n</i>	<i>MPE_n</i>	<i>MAE_n</i>	<i>MAPE_n</i>	<i>RMSE_n</i>

Tabella 3.1: Esempio di cruscotto per il monitoraggio del forecast

Nelle colonne della tabella si trovano le varie misure di accuratezza dei modelli, mentre sulle righe si trovano le quantità che devono essere previste. Sulla prima riga vengono riportati i target error, ovvero le soglie limite consentite scelte dai Demand Planner in base agli obiettivi aziendali di performance, per le varie misure di forecast error.

Se un errore sintetico di previsione tra quelli proposti nella tabella, misurato nell'intervallo di test, supera il valore del rispettivo target error, gli analisti possono decidere se correggerlo manualmente oppure di trascurarlo.

Di solito questa scelta è ponderata da alcuni specifici indicatori di performance, come per esempio la redditività o il fatturato dello specifico SKU, scelti accuratamente dai Demand Planner.

Tuttavia le metriche di valutazione dell'errore di previsione risultano spesso discordi nel quantificare la qualità delle previsioni. Infatti può accadere che per uno stesso SKU una metrica A proponga il modello 1 come migliore modello di sales forecasting, mentre la metrica B assegni un miglior valore di forecast accuracy al modello 2.

Per risolvere questo problema i Demand Planner, considerando il contesto aziendale in cui operano, costruiscono un indicatore univoco delle varie misure di errore attraverso una media ponderata. Tale indicatore viene definito come:

$$W(e_{TOT}) = \sum_{k \in K} \alpha_k * f_k(e_{tot}) \quad (3.14)$$

dove i coefficienti α_k rappresentano i pesi delle diverse misure di errore $f_k(e_{tot})$ appartenenti all'insieme K che contiene gli indicatori di errore.

Si deve tenere presente che, per esempio, non bisogna cadere nella tentazione di sommare linearmente un indicatore assoluto con una misure di errore percentuale perché ciò risulterebbe concettualmente sbagliato.

4. Times Series Forecasting

Grazie agli algoritmi di *sales forecasting* viene generata la previsione della domanda commerciale, nel periodo oggetto di interesse, sia esso regolare o promozionale.

In tutti e due i casi grazie all'analisi della serie storica baseline D_t^B , ricevuta in input dopo essere stata generata dal processo di *sales cleaning*, questi algoritmi ricavano la proiezione futura delle vendite \hat{y}_t in funzione dell'andamento della domanda registrata nel passato.

Va evidenziato il fatto che le previsioni \hat{y}_t prodotte dai modelli statistici riguardano esclusivamente periodi di vendita regolari (poiché la baseline D_t^B presenta solo vendite regolari), mentre se si è interessati a previsioni per periodi promozionali si deve aggiungere alla previsione \hat{y}_t l'impatto promozionale stimato $\widehat{\Delta}_t$ attraverso il metodo che si è utilizzato per il calcolo di quest'ultimo nel processo di sales cleaning. In altre parole si utilizza la seguente formula:

$$\hat{y}_t = \begin{cases} \hat{y}_t & \text{se } t \neq \text{promozione} \\ \hat{y}_t * \widehat{\Delta}_t & \text{se } t = \text{promozione} \end{cases} \quad (4.1)$$

Come si è discusso nel capitolo precedente la qualità predittiva dei modelli di *time series forecasting* è misurata attraverso il calcolo degli errori di previsione in corrispondenza dell'intervallo di verifica⁴, avendo precedentemente addestrato i modelli sull'analisi dei valori storici disponibili sull'intervallo di stima⁵.

In generale il modello che minimizza una determinata metrica o la combinazione di più metriche (scelte accuratamente dai Demand Planner in base al contesto aziendale) si rivela più accurato nella previsione della domanda e quindi viene adottato per la generazione della previsione oggetto di interesse.

4.1 Fasi del processo di sales forecasting

Seguendo le indicazioni di Milanato (2008), data una serie storica di domanda base D_t^B le fasi del processo di sales forecasting in generale si riassumono nei seguenti passi:

- 1) Analisi delle caratteristiche quantitative della serie storica (sales cleaning)

⁴ Quando si parla di Intervallo di verifica si fa riferimento indifferentemente a uno degli intervalli di verifica individuati con uno dei tre metodi descritti nel capitolo 3.

⁵ Anche in questo caso si intende un intervallo di stima riferendosi indifferentemente a uno degli intervalli di stima analizzati dai tre metodi descritti nel capitolo 3.

- 2) Selezione delle famiglie di algoritmi candidate a generare la previsione futura
- 3) Best-fit parametrico per la determinazione dei valori ottimali per i parametri degli algoritmi di sales forecasting selezionati
- 4) Best-fit algoritmico per la selezione del modello ottimale
- 5) Generazione del sales forecast utilizzando l'algoritmo migliore selezionato
- 6) Valutazione dell'errore di previsione in presenza di nuovi valori storici
- 7) Aggiornamento dei parametri degli algoritmi (ritorno ai punti 1,2 e 3)

Fase 1. La serie storica di domanda originale D_t viene inizialmente analizzata tramite il processo di *sales cleaning* (descritto approfonditamente nel capitolo 2) dove vengono rettificati gli outliers, successivamente vengono studiate le caratteristiche quantitative, quali le componenti di regolarità (trend e/o stagionalità) e viene individuata la tipologia di serie storica (continuativa o sporadica).

Fase 2. A seconda della tipologia di serie storica analizzata, viene selezionato un insieme di algoritmi candidati alla generazione del sales forecast. Per esempio se la serie è di tipo sporadico intermittente, ovvero presenta più del 25% di zeri, allora il modello Croston è un papabile candidato alla previsione di domanda di quella specifica serie.

Fase 3. Gli algoritmi selezionati nella fase 2 vengono addestrati sui valori disponibili per la serie storica nell'intervallo di stima. Successivamente i parametri numerici propri di ciascun algoritmo vengono settati per fare in modo di minimizzare l'errore di previsione calcolato da quel determinato algoritmo per predire i valori di domanda dell'intervallo di verifica. Viene selezionata quindi la combinazione di parametri, per ogni algoritmo, che minimizza una determinata metrica di errore scelta tra: ME, MPE, MAE, MAPE e RMSE tutte descritte dettagliatamente nel capitolo 3.

Fase 4. Gli algoritmi vengono quindi confrontati tra loro per selezionarne il migliore, ovvero quello che presenta la migliore *forecast accuracy*, limitatamente ai valori storici presenti nell'intervallo di test. In particolare viene scelto l'algoritmo che ha generato il minor errore di previsione. Questa scelta implica il fatto che si suppone che l'algoritmo selezionato sia il migliore anche per la stima dei valori futuri di domanda.

Fase 5. Una volta scelto, l'algoritmo di sales forecasting viene applicato sia lungo l'orizzonte di pianificazione della domanda, per la generazione del piano previsionale, sia lungo l'intervallo di verifica per calcolare e tenere conto dell'errore di previsione compiuto nei periodi passati.

Fase 6. Nel momento in cui si applica il modello di *sales forecasting* selezionato su tutta la serie storica di domanda disponibile, ovvero per le osservazioni temporali $t = 1, \dots, T$, si misura lo scostamento tra l'ultima previsione effettuata dall'algorithm selezionato, ovvero per $t = T$, e il corrispondente valore della serie di domanda. Possono accadere due diverse situazioni:

1. l'errore di previsione, misurato all'istante T , è superiore ad un valore soglia massimo ammesso (target error). A causa di questo, l'intera serie storica di domanda viene riprocessata interamente (time series analysis, selezione degli algoritmi, best fit) con la conseguenza che il sistema potrebbe selezionare come migliore un altro algoritmo ed anche un nuovo set di parametri ottimali da applicare alla serie storica;
2. l'errore di previsione è inferiore al target error, che identifica una previsione poco accurata. In questo caso l'algorithm è ritenuto valido per la generazione della previsione nell'orizzonte temporale pianificato, e per questo viene riutilizzato senza alcuna modifica dei parametri interni precedentemente identificati.

Va sottolineato il fatto che se esiste un periodo congelato di domanda (ad esempio l'obiettivo è prevedere i prossimi tre periodi $t + 1, t + 2, t + 3$, allora le considerazioni fatte per il precedente ciclo di sales forecasting non vengono rimesse in discussione. Ciò comporta che nel nuovo piano di sales forecasting venga utilizzato lo stesso modello scelto nel precedente piano.

Lo svantaggio principale del metodo del *refresh*, ovvero della rigenerazione completa del piano di sales forecasting, rispetto al caso di rigenerazione per eccezione (*update*) caratterizzato da insufficiente accuratezza predittiva dei modelli, è quello di essere computazionalmente più oneroso. Tuttavia il metodo del *refresh* è il miglior approccio utilizzabile nel caso in cui sia disponibile un nuovo periodo. Infatti bisogna tenere conto che un algoritmo rivelatosi il migliore per predire la domanda dei periodi $t+1, t+2, t+3, \dots$ disponendo all'istante corrente t dei T valori storici $t=1, \dots, T$, non necessariamente risulta ottimale anche per la futura predizione della domanda nei periodi $t+2, t+3, t+4, \dots$, disponendo all'istante corrente $t+1$ dei T valori storici $t=2, \dots, T+1$ (Milanato, 2008).

Fase 7. A seconda di quale dei due approcci viene utilizzato per il riaddestramento dei modelli al procedere del tempo (modalità refresh o update), i parametri dei nuovi modelli vengono ritirati, seguendo lo schema del best fit parametrico, e le nuove serie storiche aggiornate vengono rianalizzate e depurate dagli eventuali outliers, mediante l'impiego di tecniche di time series analysis.

4.2 Classificazione dei modelli di time Series Forecasting

Esistono varie classificazioni dei modelli di Time Series Forecasting. Tuttavia una caratteristica comune a tutti i modelli è quella che la storia passata sia in qualche misura esplicativa del comportamento futuro della domanda commerciale. I modelli di sales forecasting possono essere classificati sulla base della:

1. **tipologia di domanda commerciale:** *modelli per serie continue o modelli per serie sporadiche.*

Esempi di modelli per serie continue, ovvero a domanda regolare, sono i seguenti: medie mobili, smorzamenti esponenziali, modello di Holt-Winters, scomposizione della serie, modelli autoregressivi, regressione lineare. Mentre i modelli per serie sporadiche (a domanda irregolare) caratterizzati da una densità elevata di zeri presenti nella serie storica, sono ad esempio il modello di Croston, il modello di Poisson o il modello di Syntetos-Boylan;

2. **tipologia di prodotti:** *prodotti a domanda continua nel tempo o prodotti di nuova introduzione (new product forecasting).*

Tutti i modelli citati per la tipologia di domanda commerciale sono esempi di modelli per prodotti a domanda continua nel tempo. Mentre modelli per prodotti di nuova introduzione sono i seguenti: modelli basati su analogia, curve di regressione logistica, applicazione di pattern percentuali, modelli di sales budgeting (approcci elementari di ripartizione di quantità di domanda aggregate su livelli gerarchici di prodotto, mercato, tempo);

3. **tipologia di variabili:** *input dipendente o input indipendente.*

Tutti i modelli che fanno parte della tipologia di domanda commerciale, sono caratterizzati dal fatto di avere come input dipendente la domanda commerciale stessa, rilevata a consuntivo nei passati periodi di vendita. Se vengono considerate come variabili indipendenti dei fattori esogeni (non controllabili da decisioni aziendali) o endogeni (driver di business controllabili dell'azienda), si possono utilizzare dei modelli esplicativi di regressione lineare.

Un altro tipo di classificazione dei modelli di sales forecasting riguarda **l'approccio utilizzato**, ovvero:

- *metodo informale*, dove il metodo naïve è il metodo più utilizzato per la sua semplicità;
- *metodo delle serie storiche*, che a sua volta si suddivide in *approccio classico*, per esempio con il modello di lisciamento esponenziale

semplice e il modello di Holt-Winters (HW), e in *approccio moderno* con il modello di Croston, i modelli White Noise, MA, AR, ARMA, ARIMA e SARIMA;

- *modelli di regressione* come il modello lineare;
- *modelli di previsione avanzati*: reti neurali;
- *new product forecasting*.

Ognuno di questi metodi presenta vantaggi e svantaggi. Ciascuno di essi può essere applicato a casi diversi, o allo stesso caso, al fine di disporre di previsioni alternative.

Va sottolineato il fatto che in letteratura sono presenti molti altri metodi o varianti di quelli già citati.

4.3 Metodo informale: Naïve Forecasting

Il modo più semplice per generare una previsione di domanda \widehat{y}_{t+1} essendo a conoscenza della serie di domanda originale y_t è quello di uguagliare tale previsione all'ultimo valore disponibile della serie di domanda, ovvero: $\widehat{y}_{t+1} = y_t \equiv y_{t-1}$

In particolare se l'obiettivo riguarda una previsione per periodi promozionali allora il forecast sarà uguagliato alle vendite del più recente evento promozionale (di tipologia simile a quello che deve essere previsto), presente nella serie storica di domanda y_t (Trusov,2006).

Questo metodo, chiamato *Naïve Forecasting*, è privo di fondamenti matematico/statici ma a causa della sua semplicità, sia concettuale che di calcolo, viene utilizzato come benchmark con cui confrontare i modelli di sales forecasting. Infatti è auspicabile investire tempo e risorse computazionali per la taratura dei modelli complessi, come Holt-Winters, solo quando l'accuratezza predittiva risulta significativamente superiore a quella ottenuta con il metodo naïve (Milanato, 2008).

4.4 Metodo delle serie storiche

Utilizzando il metodo delle serie storiche per descrivere il processo generatore dei dati di una serie storica $\{y_t\}_{t=1}^T$ relativa ad una variabile Y si è soliti usare un modello stocastico generale come il seguente:

$$Y_t = f(t) + u_t \quad (4.2)$$

dove $f(t)$ è una sequenza completamente deterministica che costituisce la parte sistematica della serie; mentre u_t è una sequenza di variabili casuali (da qui in avanti chiamate v.c.) che rappresenta la parte stocastica della serie ed obbedisce ad una determinata legge di probabilità. Va evidenziato che

entrambe le componenti non sono individualmente osservabili, ma vanno determinate sulla base del campione.

Da quanto visto nel paragrafo 4.2 si utilizzano due tipi di approcci di analisi delle serie storiche per trattare il modello (4.2) ovvero l'*Approccio classico* o l'*Approccio moderno* (vedasi paragrafo 4.4.4).

Nell'**Approccio classico** (o tradizionale) si suppone che “esista una legge di evoluzione temporale del fenomeno, rappresentata da $f(t)$ ”; mentre “la componente casuale u_t viene invece assunta a rappresentare l'insieme delle circostanze, ciascuna di entità variabile, che non si vogliono o non si possono considerare esplicitamente in Y_t . I residui di Y_t non spiegati da $f(t)$, vengono pertanto imputati al caso ed assimilati ad errori accidentali” (Di Fonzo e Lisi, 2005: p.29-30).

Secondo un punto di vista statistico, seguendo l'approccio classico, viene ipotizzato che la componente stocastica del modello (4.2) sia generata da un processo *white noise*, ovvero da una successione di v.c. iid (indipendenti ed identicamente distribuite), di media nulla e varianza costante; in formule risulta:

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad \text{dove} \quad \begin{aligned} E[\varepsilon_t] &= 0 & \forall t \\ \text{VAR}[\varepsilon_t] &= \sigma_\varepsilon^2 & \forall t \\ E[\varepsilon_r \varepsilon_s] &= 0 & \forall r, s; \quad r \neq s \end{aligned} \quad (4.3)$$

“Una successione di v.c. u_t per cui si abbia $\text{COV}[u_r, u_s] = 0, \forall r, s; \quad r \neq s$ indipendentemente dal fatto che sia o meno un *white noise*, viene detto processo stocastico a componenti incorrelate. Per contro, quando $\text{COV}[u_r, u_s] \neq 0$, per qualche $r \neq s$, si ha un processo a componenti correlate” (Di Fonzo e Lisi, 2005: p.30).

Nei successivi paragrafi seguendo l'approccio classico per l'analisi delle serie storiche vengono descritti nell'ordine i modelli di lisciamiento esponenziale semplice (paragrafo 4.4.1), il modello di Holt-Winters (paragrafo 4.4.2) e il modello di Croston (paragrafo 4.4.3).

Tuttavia va evidenziato fin da subito che i previsori proposti nel seguito (ovvero costruiti utilizzando l'approccio classico) non sono basati su alcun modello probabilistico. A maggior ragione “poiché non viene richiesta alcuna fase formale di identificazione non è neppure corretto parlare di modello ma piuttosto di metodi ad hoc. Proprio per questi motivi, i previsori basati sul lisciamiento esponenziale non godono, in generale, di particolari proprietà a priori né permettono di formulare affermazioni probabilistiche sulle previsioni” (Di Fonzo e Lisi, 2005).

4.4.1 Modello di lisciamiento esponenziale semplice

I modelli di *lisciamiento esponenziale* (detti *exponential smoothing*), considerando l'intero orizzonte di dati storici disponibili, assegnano un peso decrescente in modo esponenziale ai valori della domanda dal momento che quest'ultimi sono più recenti e quindi assumono maggior importanza rispetto ai valori più arretrati.

Il modello più comunemente utilizzato è il *modello di Brown (1956)*, chiamato anche *lisciamiento esponenziale semplice*. Esso genera la previsione \widehat{y}_{t+1} per il periodo $t + 1$ in modo ricorsivo considerando sia la domanda registrata nell'ultimo periodo storico y_t sia l'ultima previsione \widehat{y}_t generata all'istante $t - 1$, per il periodo t nel seguente modo:

$$\widehat{y}_{t+1} = \alpha * y_t + (1 - \alpha) * \widehat{y}_t \quad (4.4)$$

dove α è un coefficiente di ponderazione e lisciamiento dei due termini di domanda e sales forecast relativi al periodo t . Va evidenziato il fatto che al crescere del peso α aumenta la rilevanza attribuita dal modello all'ultimo valore storico di domanda. Ciò significa che la previsione è reattiva nei confronti di variazioni improvvise nelle vendite.

Sviluppando la (4.4) sostituendo \widehat{y}_t con la funzione della domanda storica del periodo $t - 1$ si ottiene:

$$\widehat{y}_{t+1} = \alpha * y_t + (1 - \alpha) * \widehat{y}_t = \alpha * y_t + (1 - \alpha) * \{ \alpha * y_{t-1} + (1 - \alpha) * \widehat{y}_{t-1} \} \quad (4.5)$$

Precedendo a ritroso anche per i periodi $t - 2, t - 3, \dots$ si ottiene una relazione di dipendenza della previsione della domanda \widehat{y}_{t+1} solo per i valori storici della domanda:

$$\widehat{y}_{t+1} = \alpha * y_t + (1 - \alpha) * y_{t-1} + \alpha * (1 - \alpha)^2 * y_{t-2} + \alpha * (1 - \alpha)^3 * y_{t-3} + \dots \quad (4.6)$$

La (4.6) si può esprimere sinteticamente come:

$$\widehat{y}_{t+1} = \alpha \sum_{k=0}^{T-1} (1 - \alpha)^k * y_{t-k} \quad (4.7)$$

Va evidenziato che se il parametro $\alpha \approx 0$, viene attribuito un peso uniforme a tutti i periodi storici considerati nel calcolo della previsione \widehat{y}_{t+1} , mentre se $\alpha \approx 1$ il modello si adatta quasi perfettamente alle variazioni casuali nei valori più recenti della domanda.

Nel caso in cui $\alpha = 0$ il modello di lisciamiento esponenziale semplice coincide con il metodo naïve.

Per l'algoritmo di lisciamiento esponenziale semplice, il best-fit parametrico consiste nel determinare il valore del parametro α contenuto nell'intervallo

[0;1], dove un'opportuna metrica di errore sia minima nell'intervallo di test della serie storica oggetto di analisi, ovvero:

$$\alpha | 0 \leq \alpha \leq 1 \cup f_{\alpha}(E_{TOT}) = \min_k \{ f_k(E_{TOT}) \} \quad (4.8)$$

essendo $f_k(E_{TOT})$ il valore di una metrica di errore f registrata in corrispondenza del parametro k assegnato al coefficiente di liscio per il modello di Brown.

4.4.2 Modello di Holt-Winters

Il modello di Holt-Winters è l'algoritmo "più utilizzato per calcolare la previsione di domanda per serie storiche continue, caratterizzate dalla presenza di trend e stagionalità. Se le serie storiche in input risultano sufficientemente regolari, il sales forecast ottenibile impiegando tale algoritmo ricorsivo risulta essere di qualità elevata" (Milanato, 2008).

Vengono considerate tre componenti di regolarità all'interno di una serie storica continua:

- componente media (α) : già considerata nel modello di Brown
- componente tendenziale (β): inclusa nel modello di Holt (1957)
- componente stagionale (γ): inclusa nel modello di Winters (1960)

I tre coefficienti numerici, inclusi nell'intervallo [0,1], definiscono l'intensità di liscio. In particolare:

- valori di $\alpha \approx 1$ indicano che il modello si adatta quasi perfettamente alle variazioni casuali nei valori più recenti della domanda;
- valori di $\beta \approx 1$ indicano una rilevante presenza di trend lineare, crescente o decrescente nel tempo;
- valori di $\gamma \approx 1$ indicano una significativa presenza della componente di stagionalità, oscillante con regolarità periodica nel tempo.

Nel **modello di Holt**, chiamato anche *liscio esponenziale doppio* che in inglese viene definito *double exponential smoothing*, vengono considerate le sole componenti di media e trend nell'analisi delle caratteristiche quantitative di una serie storica. Per questa ragione non è possibile applicare questo modello a serie storiche con presenza di ciclicità stagionale, a meno che queste non siano prima destagionalizzate con il processo di *sales cleaning* descritto nel capitolo 2.

Le componenti di media e di tendenza vengono stimate in modo separato utilizzando le seguenti formule ricorsive:

- componente media (M_t):

$$M_t = \alpha * y_t + (1 - \alpha) * (M_{t-1} + T_{t-1}) \quad (4.9)$$

- componente di tendenza lineare (T_t):

$$T_t = \beta * (M_t - M_{t-1}) + (1 - \beta) * T_{t-1} \quad (4.10)$$

dove la differenza $M_t - M_{t-1}$ indica la variazione con segno della domanda media fra due periodi storici consecutivi.

Le (4.9) e (4.10) vengono applicate cronologicamente alla serie storica di domanda, ovvero partendo dal primo periodo $t = 1$ fino all'ultimo periodo passato T . Convenzionalmente si assegnano dei valori iniziali M_0 e T_0 affinché le quantità M_1 e T_1 possano essere calcolate.

La previsione della domanda $\widehat{y_{t+k}}$ per i periodi futuri dell'intervallo di previsione $t + k$, $k = 1, 2, 3, \dots$, si ottiene come:

$$\widehat{y_{t+k}} = M_t + k * T_t \quad (4.11)$$

dove M_t e T_t sono i valori calcolati al termine dell'ultimo periodo storico $T = t$ in modo ricorsivo partendo dal primo valore di domanda disponibile nella serie storica.

Alla domanda base dello SKU per ciascun periodo di previsione può accadere, a volte, di aggiungere un pattern di vendita non lineare, modellato attorno all'incremento costante di vendita T_t , di durata largamente inferiore rispetto all'intero orizzonte di pianificazione. Questo è quello che succede, per esempio, quando si è in presenza di un evento promozionale. Quando succede ciò, il modello di Holt, descritto in precedenza, può essere modificato al fine di includere nella previsione della domanda la presenza di una componente di tendenza non lineare (propria della promozione) che non può essere altresì registrata nel corso dell'analisi della serie storica oggetto di studio. Si parlerà quindi del **modello di Holt a tendenza smorzata**.

La (4.11) può essere modificata per generare la nuova classe di modelli esponenziali a tendenza smorzata, nel seguente modo:

$$\widehat{y_{t+k}} = M_t + \delta_k * T_t \quad (4.12)$$

dove δ_k è la serie dei coefficienti di riduzione o incremento della componente di tendenza lineare T_t .

In generale la (4.9) e la (4.10) rispettivamente utilizzate per il calcolo delle componenti di media e tendenza del modello di Holt (versione base) diventano:

$$M_t = \alpha * y_t + (1 - \alpha) * (M_{t-1} + T_{t-1}) * \delta_t \quad (4.13)$$

$$T_t = \beta * (M_t - M_{t-1}) + (1 - \beta) * T_{t-1} * \delta_t \quad (4.14)$$

dove i coefficienti δ_t sono un ulteriore liscio delle rispettive componenti calcolate progressivamente per ogni periodo $t = 1, \dots, T$.

Se oltre alle componenti di media (modello di Brown) e di tendenza (modello di Holt) si vuole tenere conto anche della componente regolare di ciclicità dovuta alle stagioni di vendita, si parlerà del **modello di Winters**, chiamato anche liscio esponenziale triplo (che in inglese viene detto triple exponential smoothing).

Le componenti di media, trend e stagionalità sono stimate secondo le seguenti formule ricorsive:

- componente media (M_t):

$$M_t = \alpha * \frac{y_t}{S_{t-L}} + (1 - \alpha) * (M_{t-1} + T_{t-1}) \quad (4.15)$$

- componente di tendenza lineare (T_t):

$$T_t = \beta * (M_t - M_{t-1}) + (1 - \beta) * T_{t-1} \quad (4.16)$$

- componente stagionale (S_t):

$$S_t = \gamma * \frac{y_t}{M_t} + (1 - \gamma) * S_{t-L} \quad (4.17)$$

dove S_{t-L} è il coefficiente adimensionale mediante il quale la serie storica di domanda y_t viene destagionalizzata. È la componente stagionale, calcolata nel periodo $t - L$, distante dal periodo corrente un intero ciclo stagionale di lunghezza pari a L (ad esempio $L=12$ mesi o 52 settimane per stagionalità annuali). Per ipotesi la lunghezza L del ciclo stagionale di domanda, espressa nelle opportune unità di misura temporali, è un parametro noto del modello di Holt-Winters.

Va evidenziato che la (4.15), la (4.16) e la (4.17) vengono applicate in senso cronologico, partendo dal primo periodo storico disponibile $t = 1$ fino all'ultimo periodo passato T . In analogia al modello di Holt, convenzionalmente si assegnano dei valori iniziali M_0 e T_0 affinché le quantità M_1 e T_1 possano essere calcolate.

Dall'analisi delle (4.15), (4.16) e (4.17) si sottolinea il fatto che:

- la quantità $\frac{y_t}{S_{t-L}}$ rappresenta la domanda base del periodo t , depurata dalla componente stagionale che ne altera il valore rispetto al valor medio della domanda destagionalizzata e detrendizzata;

- la quantità $\frac{y_t}{M_t}$ si riferisce alla domanda base non smorzata dal coefficiente di stagionalità per il periodo t , divisa per il valore medio M_t . “La successiva applicazione dei coefficienti stagionali S_t ai valori medi M_t in modalità previsiva genera la modulazione della domanda futura media rispetto ai pattern periodici propri dei cicli stagionali di vendita” (Milanato, 2008).

La previsione della domanda $\widehat{y_{t+k}}$ per i periodi futuri dell’intervallo di previsione $t + k$, $k = 1, 2, 3, \dots$, si ottiene come:

$$\widehat{y_{t+k}} = (M_t + k * T_t) * S_{t+k-L} \quad (4.18)$$

dove M_t , T_t e S_{t-L} sono i valori calcolati al termine dell’ultimo periodo storico $T = t$ in modo ricorsivo partendo dal primo valore di domanda disponibile nella serie storica.

Strutturalmente, il modello appena presentato, noto come *modello moltiplicativo congiunto di Holt-Winters*, rappresenta:

- la componente di tendenza T_t in modo additivo (modello di Holt);
- la componente di stagionalità S_t in modo moltiplicativo (modello di Winters).

Va sottolineato infine il fatto che “condizione necessaria e sufficiente affinché il modello di Holt-Winters sia correttamente applicabile a serie storiche continue e regolari, affette da tendenza lineare e stagionalità, è di disporre di almeno $2L$ periodi di domanda nella serie storica di vendita nell’intervallo di training dell’algoritmo, pari cioè a due cicli completi di vendita” (Milanato, 2008).

Questo vincolo è *necessario* in quanto il modello non riesce a generare i coefficienti stagionali S_{t-L} , che vengono identificati dal confronto fra coppie di valori di domanda base y_t equidistanti un ciclo di vendita pari a L . Tuttavia se il modello viene utilizzato per prevedere i valori di domanda all’interno del test set, il vincolo precedentemente descritto è una condizione sufficiente.

Finora è stato mostrato come la (4.15), la (4.16) e la (4.17) del modello di Holt-Winters vengano calcolate in modo ricorsivo partendo dal periodo $t = 1$ della serie storica fino all’ultimo T .

Tuttavia affinché si riescano a calcolare le quantità M_1, T_1 e S_1 occorre determinare i valori iniziali dei parametri M_0, T_0 e un intero ciclo di valori di stagionalità pari a S_t .

Per fare ciò si utilizzano le seguenti formule, calcolate nell'ordine presentato:

- *valore iniziale della componente di trend*, è la differenza tra i valori di domanda per i primi due cicli stagionali divisi per la lunghezza L pari a un ciclo stagionale:

$$T_0 = \frac{1}{L} * \left(\frac{1}{L} * \sum_{t=L+1}^{2L} y_t - \frac{1}{L} * \sum_{t=1}^L y_t \right) \quad (4.19)$$

- *valore iniziale della componente media*, è la somma fra la domanda media del secondo ciclo stagionale e la componente iniziale di tendenza moltiplicata per la metà di un ciclo stagionale:

$$M_0 = \frac{1}{L} * \sum_{t=L+1}^{2L} y_t + \frac{L}{2} * T_0 \quad (4.20)$$

- *valore iniziale dei periodi omologhi in un ciclo per la componente stagionale* per ciascun periodo $l = 1, \dots, L$ è definita come la media aritmetica del rapporto tra le domande di due periodi omologhi l successivi (ad esempio i mesi di Gennaio 2011 calcolato nel primo ciclo storico, ovvero $c=1$, e Gennaio 2012 calcolato nel secondo ciclo storico, cioè $c=2$):

$$S_{0,l} = \frac{1}{2} * \left(\frac{y_l(c=1)}{\frac{1}{L} * \sum_{t=1}^L y_t} + \frac{y_l(c=2)}{\frac{1}{L} * \sum_{t=L+1}^{2L} y_t} \right) \quad (4.21)$$

4.4.3 Modello di Croston

Il modello di Croston (formulato nel 1972), è uno tra i metodi migliori per modellare la domanda commerciale di tipo sporadico casuale⁶ in quanto i metodi classici di liscio tendono a fornire previsioni distorte (al rialzo); ed è composto essenzialmente di due fasi:

- analisi delle quantità di domanda;
- analisi del profilo temporale della domanda.

Fase 1. I valori negli intervalli di stima e verifica, vengono filtrati in modo da selezionare solo quelli non nulli per ricavare una serie storica compatta di n_V valori consecutivi $\{y_1, y_2, \dots, y_j, \dots, y_{n_V}\}$ che presenterà un coefficiente di

⁶ Le serie sporadiche casuali sono tipiche di prodotti basso vendenti, come ad esempio articoli di abbigliamento fuori moda. La difficoltà nel formulare una previsione di domanda sta, in primo luogo, nel fatto della difficile prevedibilità del prossimo periodo futuro in cui si manifesterà una domanda non nulla; e in secondo luogo all'imprevedibilità della domanda nel primo periodo futuro a domanda non nulla

variazione⁷ (CV_D) in linea di massima elevato. Esso viene utilizzato nel calcolo del primo valore futuro non nullo di domanda \hat{y}_0 , da collocare temporalmente in un periodo appartenente all'orizzonte di pianificazione, ancora da definire. Il valore \hat{y}_0 viene quindi determinato con modelli di *time series forecasting* quali:

- modelli a media mobile (vedasi Di Fonzo e Lisi, 2005 oppure Milanato, 2008)
- modelli a lisciamento esponenziale semplice (vedasi 4.4.1)
- modello base di Holt oppure il modello di Holt a tendenza smorzata (vedasi 4.4.2)

Va sottolineato il fatto che l'utilizzo del modello di Holt-Winters non risulta appropriato a causa della componente stagionale periodica di domanda che in una serie storica sporadica casuale è difficilmente riscontrabile (Milanato, 2008: pag. 451).

Fase 2. I periodi della serie storica y_t , inclusi negli intervalli di stima e verifica, vengono classificati in due tipologie in base al tipo di domanda, sia essa nulla o a valori positivi. Ovvero si considera il modello seguente:

$$y_t = x_t * z_t + \varepsilon_t \quad (4.22)$$

dove y_t è il valore di domanda osservato, z_t è il valore di domanda non nullo, x_t è una variabile casuale di Bernoulli di parametro $\pi \in [0,1]$ ed ε è un errore casuale.

Un algoritmo di conteggio calcola la distanza fra tutte le coppie di periodi successivi a domanda non nulla, partendo da $t=1$ fino a $t=T$, determinando la serie delle lunghezze Δt degli intervalli di interarrivo della domanda. Per esempio in $\{0,0,0,13,0,0,0,35,0,0,0,0,0,46,0,0,32,0,15\}$, la serie degli intervalli di interarrivo, risulta essere $\Delta t = \{4,4,6,3,2\}$. Infatti il primo valore non nullo 13 è nel periodo $t=4$ mentre il secondo valore non nullo 35 si trova nel periodo $t=8$, quindi $\Delta t = 8 - 4 = 4$.

⁷ Il coefficiente di variazione è definito dal rapporto fra deviazione standard σ_D e il valor medio μ_D della distribuzione dei valori, misurati negli intervalli di training e test set. In formule risulta:

$$CV_D = \frac{\sigma_D}{\mu_D} = \frac{\sqrt{T * \sum_{t=1}^T (y_t - \frac{1}{T} * \sum_{t=1}^T y_t)^2}}{\sum_{t=1}^T y_t}$$

Questo coefficiente esprime la percentuale di variazione della serie storica rispetto al valor medio e aumenta proporzionalmente all'aumentare della dispersione dei valori σ_D

Anche in questo caso per calcolare il primo valore previsionale $\widehat{\Delta t}_0$ si utilizzano tecniche di *time series forecasting*. Risultano appropriati modelli di media mobile e liscio esponenziale (Milanato, 2008).

Dopo aver calcolato i primi valori dei parametri \widehat{y}_0 e $\widehat{\Delta t}_0$, esistono diversi modi per utilizzare questi valori nel generare il sales forecast per il modello di Croston.

Il più semplice consiste nell'applicare in modo periodico il valore costante di domanda \widehat{y}_0 ai periodi futuri, intervallati da una distanza pari a $\widehat{\Delta t}_0$. Per esempio, se $\widehat{\Delta t}_0 = 5$ e $\widehat{y}_0 = 34$ la previsione di domanda sarà: $\{34,0,0,0,0,34,0,0,0,0,34,0,0,0,0,34, \dots\}$.

Va evidenziato come in questo caso il rapporto $\frac{\widehat{y}_0}{\widehat{\Delta t}_0}$ definisce la domanda media all'interno di due periodi consecutivi a domanda non nulla.

Per completezza espositiva si vuole sottolineare il fatto che il modello di Croston, fin qui presentato nella sua forma base, può essere modificato grazie a numerose alternative, implementate al fine di migliorare la modellazione della sporadicità intermittente della domanda, con particolare focus all'irregolarità dove sono presenti intervalli di domanda nulla (vedasi Milanato, 2008).

4.4.4 Processi stocastici: WN, AR, MA, ARMA, ARIMA e SARIMA

Nell' **Approccio moderno** dell'analisi delle serie storiche si considera la (4.2) ipotizzando che $f(t)$ manchi o sia già stata eliminata (mediante stima o altri metodi). Per questo l'attenzione viene posta quindi sulla componente stocastica u_t che si ipotizza essere un processo a componenti correlate trattabile con opportune tecniche statistiche, per estrarre qualche tipo di informazione. L'espressione della componente stocastica è la seguente:

$$u_t = g(Y_{t-1}, Y_{t-2}, \dots, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) + \varepsilon_t \quad (4.23)$$

Per questo tipo di approccio $\{Y_t\}$ è un processo stocastico⁸, ovvero una collezione di variabili casuali indicizzate dal tempo:

$$\{Y_t, t = t_1, t_2, \dots\} \quad (4.24)$$

“Per descrivere in modo completo un processo stocastico si dovrebbe specificare la distribuzione di probabilità congiunta di $(Y_{t_1}, \dots, Y_{t_T})$ per ogni insieme di istanti t_1, \dots, t_T ed ogni valore di T. Ciò è piuttosto complicato ed in genere si preferisce descrivere il processo stocastico sulla base dei momenti, e in particolare, dei primi due, delle v.c. Y_t ” (Di Fonzo e Lisi, 2005).

⁸ Per una trattazione completa si veda ad esempio Brockwell e Davis (1991)

Al variare di t media, varianza e covarianza definiscono le seguenti funzioni a valori reali:

- Funzione Media:

$$\mu_t = E[Y_t] \quad (4.25)$$

- Funzione Varianza:

$$\sigma_t^2 = VAR[Y_t] = E[Y_t - \mu_t]^2 \quad (4.26)$$

- Funzione Autocovarianza:

$$\gamma_{t_1, t_2} = E\{ [Y_{t_1} - \mu_{t_1}] [Y_{t_2} - \mu_{t_2}] \} \quad (4.27)$$

Va sottolineato il fatto che se $t_1 = t_2 = 0$ allora $\gamma_{t_1, t_2} = \sigma_t^2$

“L’autocovarianza non è altro che la covarianza tra v.c. del medesimo processo stocastico spaziate fra loro di uno sfasamento temporale pari a $k = |t_2 - t_1|$ ” (Di Fonzo e Lisi, 2005:p.162).

Per rendere più facile l’interpretazione della autocovarianza si è soliti introdurre la funzione di autocorrelazione (generalmente indicata con ACF, da AutoCorrelation Function) che non è altro che la seguente normalizzazione:

$$\rho_{t_1, t_2} = \frac{\gamma_{t_1, t_2}}{\sigma_{t_1} \sigma_{t_2}} \quad (4.28)$$

La (4.28) è una misura della correlazione tra Y_{t_1} e Y_{t_2} ma, a differenza della funzione di autocovarianza ha il vantaggio di essere un numero puro, ovvero è indipendente dall’unità di misura di Y_t .

In generale, “la correlazione tra due variabili può essere dovuta al fatto che esista effettivamente un legame diretto tra le variabili o al fatto che queste ultime sono correlate con una terza variabile. Nel contesto delle serie storiche, buona parte della correlazione tra Y_t e Y_{t-k} può essere dovuta alla correlazione che tali variabili hanno con $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$ ” (Di Fonzo e Lisi, 2005: pag.164).

Una possibile soluzione per tenere conto di quanto appena discusso è considerare la funzione di autocorrelazione parziale (PACF, da Partial AutoCorrelation Function), che misura l’autocorrelazione tra Y_t e Y_{t-k} al netto delle variabili intermedie. Detto in altri termini, la funzione di autocorrelazione parziale è pari alla correlazione condizionata:

$$P_k = Corr(Y_t, Y_{t-k} | Y_{t-1}, \dots, Y_{t-k+1}) \quad (4.29)$$

Processi stocastici stazionari

Seguendo Di Fonzo e Lisi (2005), si definisce un processo stocastico $\{Y_t\}$ come *stazionario del secondo ordine*, o *debolmente stazionario*, se la sua media è costante e la sua funzione di autocovarianza dipende solo da k, ovvero:

$$\begin{aligned} E[Y_t] &= \mu & \forall t \\ Cov[Y_t, Y_{t+k}] &= \gamma_k & \forall t \end{aligned} \quad (4.30)$$

Ne consegue che dato un processo stazionario⁹ $\{Y_t\}$ è possibile calcolare in modo univoco la sua funzione di autocovarianza γ_k e quindi la funzione di correlazione ρ_k . Tuttavia data una funzione di autocovarianza il processo stazionario $\{Y_t\}$, il quale possiede quella specifica funzione di autocovarianza, non è unico. Infatti affinché tale processo risulti unico, esso oltre a essere stazionario deve essere anche invertibile. Formalmente, seguendo Di Fonzo e Lisi (2005), un processo $\{Y_t\}$ è detto *invertibile* se esiste una funzione lineare $h(\cdot)$ e un processo white noise $\{\varepsilon_t\}$ tale che, per ogni t , si ha:

$$Y_t = h(Y_{t-1}, Y_{t-2}, \dots) + \varepsilon_t \quad (4.31)$$

Processo rumore bianco (white noise)

Il processo *white noise* $\{\varepsilon_t\}$ è uno dei principali processi stocastici caratterizzato da una sequenza di variabili casuali incorrelate a media e varianza costanti.

La funzione di autocorrelazione, identicamente uguale a P_k è la seguente:

$$\rho_k = \begin{cases} 1 & k = 0 \\ 0 & k = \pm 1, \pm 2, \dots \end{cases} \quad (4.32)$$

Processo a media mobile MA(q)

Dato un processo *white noise* $\{\varepsilon_t\}$ di media 0 e varianza σ_ε^2 , un processo a media mobile di ordine q (in sigle MA(q) e in inglese chiamato *Moving Average*) è definito come:

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (4.33)$$

con $\theta_j (j = 0, \dots, q)$ parametri costanti.

Utilizzando l'operatore ritardo B , tale per cui $B^h Y_t = Y_{t+h}$ dove $h = 0, 1, 2, \dots$ e indicando con

$\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ l'operatore polinomiale a media mobile, la (4.33) può essere scritta equivalentemente anche come:

$$Y_t = \theta(B) \varepsilon_t \quad (4.34)$$

⁹ Per semplicità il termine "stazionario" si riferisce al concetto di "debolmente stazionario" e non al concetto di stazionarietà in senso stretto (vedasi Di Fonzo & Lisi, 2005, pag.162-166)

La media di un processo MA(q) è pari a :

$$E[Y_t] = E[\varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}] = 0 \quad (4.35)$$

L'autocovarianza di un processo MA(q) è definita come:

$$\gamma_k = \begin{cases} \sigma_\varepsilon^2 (1 + \sum_{i=1}^q \theta_i^2) & k = 0 \\ \sigma_\varepsilon^2 (-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \dots + \theta_{q-k} \theta_q) & k = 1, \dots, q \\ 0 & k > q \end{cases} \quad (4.36)$$

La funzione di autocorrelazione risulta quindi essere:

$$\rho_k = \begin{cases} 1 & k = 0 \\ \frac{-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \dots + \theta_q^2} & k = 1, \dots, q \\ 0 & k > q \end{cases} \quad (4.37)$$

Dal momento che media, varianza e covarianza sono costanti e non dipendono dal tempo, un processo MA(q) è sempre stazionario. Per verificare l'invertibilità si utilizza l'equazione caratteristica $\theta(B) = 0$. Viene dimostrato che un processo MA(q) è invertibile se tutte le radici dell'equazione caratteristica sono in modulo maggiori di 1 (Hamilton, 1994)

Processo autoregressivo AR(p)

Dato un processo white noise $\{\varepsilon_t\}$ di media nulla e varianza σ_ε^2 ; $\{Y_t\}$ è un processo autoregressivo di ordine p, e lo si indica con AR(p) (dall'inglese AutoRegressive) se:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (4.38)$$

dove ϕ_i ($i = 0, \dots, p$) sono parametri costanti.

In modo analogo a quanto visto per il processo MA(q), utilizzando l'operatore ritardo B e indicando con $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ l'operatore polinomiale autoregressivo, un processo AR(p) può essere scritto come:

$$\phi(B)Y_t = \phi_0 + \varepsilon_t \quad (4.39)$$

La media di un processo AR(p) è pari a:

$$\mu = E[Y_t] = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p} + \varepsilon_t \quad (4.40)$$

La funzione di autocovarianza è data da:

$$\gamma_k = \begin{cases} \phi_1 \gamma_1 + \dots + \phi_p \gamma_p + \sigma_\varepsilon^2 & k = 0 \\ \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} & k > 0 \end{cases} \quad (4.41)$$

La funzione di autocorrelazione è:

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p} \quad (4.42)$$

Il processo AR(p) è sempre invertibile, mentre viene dimostrato che un processo AR(p) è stazionario se tutte le p radici dell'equazione caratteristica $\phi(B) = 0$ sono in modulo, maggiori di 1 (Wei, 1990).

Processo autoregressivo a media mobile ARMA(p,q)

Poiché i processi AR e MA possono avere un gran numero di parametri, si preferisce utilizzare una rappresentazione più parsimoniosa considerando un modello misto, proposto da Box e Jenkins (1976).

Dato un processo white noise $\{\varepsilon_t\}$ di media nulla e varianza σ_ε^2 ; $\{Y_t\}$ è un processo autoregressivo a media mobile di ordine (p,q). Questo viene indicato con la notazione ARMA(p,q) (dall'inglese AutoRegressive Moving Average) che si può scrivere come:

$$Y_t - \sum_{i=1}^p \phi_i Y_{t-i} = \phi_0 + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (4.43)$$

Va evidenziato il fatto che quando p=0 il processo ARMA(p,q) \equiv MA(q); mentre quando q=0 il processo ARMA(p,q) \equiv AR(p).

Anche in questo caso la (4.43) può essere riscritta utilizzando l'operatore ritardo B, ottenendo:

$$\phi(B)Y_t = \phi_0 + \theta(B)\varepsilon_t \quad (4.44)$$

Un processo ARMA(p,q) è stazionario se "tutte le p radici dell'equazione caratteristica $\phi(B) = 0$ sono, in modulo, maggiori di 1, mentre è invertibile se le q radici dell'equazione $\theta(B) = 0$ sono tutte, in modulo, maggiori di 1" (Di Fonzo e Lisi, 2005).

La funzione di autocovarianza fino al ritardo q dipende sia dalla parte autoregressiva, sia da quella a media mobile; mentre per $k > q$, dipende solamente dalla parte autoregressiva, ovvero:

$$\gamma_k = \begin{cases} \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} + E(Y_{t-k} \varepsilon_t) - \theta_1 E(Y_{t-k} \varepsilon_{t-1}) + \dots + \theta_q E(Y_{t-k} \varepsilon_{t-q}) & k = 0, 1, \dots, q \\ \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} & k \geq q + 1 \end{cases} \quad (4.45)$$

La funzione di autocorrelazione è pari a:

$$\rho_k = \begin{cases} \frac{\gamma_k}{\gamma_0} & k = 0, 1, \dots, q \\ \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p} & k \geq q + 1 \end{cases} \quad (4.46)$$

Processi stocastici non stazionari

Specialmente nel caso di variabili economiche, è frequente che un processo stocastico non sia stazionario. Prendendo in considerazione i primi due momenti, la non stazionarietà può dipendere dal fatto che la media e/o i momenti secondi non siano costanti. Nel primo caso si parlerà di non stazionarietà in media, mentre nel secondo di non stazionarietà in varianza.

“Alcuni tipi di stazionarietà possono essere eliminati mediante differenziazione. In particolare, si dice che un processo è non stazionario omogeneo di grado d se diventa stazionario a seguito di d differenziazioni successive” (Di Fonzo e Lisi, 2005).

Va sottolineato il fatto che non tutti i problemi legati alla non stazionarietà possono essere risolti tramite differenziazioni. Questo è il caso di processi stazionari in media ma non in varianza.¹⁰

Processo autoregressivo integrato a media mobile ARIMA(p,d,q)

I processi ARIMA, seguendo le indicazioni di Box e Jenkins (1976), possono essere estesi al caso di processi non stazionari omogenei di grado d , ovvero che possono essere resi stazionari a seguito di d differenziazioni successive.

Sia definito un processo white noise $\{\varepsilon_t\}$ di media nulla e varianza σ_ε^2 e si indichi X_t la d -esima differenza di Y_t , $X_t = (1 - B)^d Y_t$. Un processo $\{Y_t\}$ è autoregressivo integrato a media mobile di ordine (p,d,q) , e viene indicato con ARIMA(p,d,q) (dall'inglese AutoRegressive Integrated Moving Average), se $\{X_t\}$ è un processo ARMA (p,q). In formule valgono le espressioni equivalenti:

$$X_t = (1 - B)^d Y_t \quad (4.47)$$

$$X_t = \phi_0 + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (4.48)$$

Utilizzando l'operatore differenza ed i polinomi, rispettivamente, autoregressivo e a media mobile, la (4.47) e (4.48) possono essere scritte in funzione di Y_t , ottenendo la seguente relazione di sintesi:

$$\phi(B)(1 - B)^d Y_t = \phi_0 + \theta(B)\varepsilon_t \quad (4.49)$$

Va sottolineato il fatto che per:

- $d=q=0$ la classe di processi ARIMA(p,d,q) si riduce a quella dei processi AR(p)
- $p=d=0$ si ottiene la classe dei processi MA(q)
- $d=0$ si ottiene la classe dei processi ARMA(p,q)
- $p=d=q=0$ si ottiene il processo white noise (WN)

¹⁰ Vedasi Di Fonzo & Lisi 2005: pag.183

Processi stagionali

I processi presentati nei precedenti paragrafi riescono a modellare andamenti di tipo periodico di tipo stagionale introducendo nel modello un gran numero di parametri. Dall'altro canto i metodi per trattare la stagionalità, come il modello di Holt-Winters visto nel paragrafo 4.4.2, si basano sul presupposto che la componente stagionale sia deterministica e indipendente dalle altre componenti. Tuttavia, spesso accade che la stagionalità sia stocastica e correlata con le componenti stagionali, e quindi si vogliono modellare congiuntamente le componenti, piuttosto che separarle.

Processo stagionale ARIMA(p,d,q) x (P,D,Q)s detto SARIMA

Per questo motivo, Box e Jenkins (1976) hanno esteso la classe dei processi ARIMA per trattare anche un comportamento di tipo periodico (stazionario o non stazionario) alla classe dei processi ARIMA stagionali, detti anche SARIMA (Seasonal ARIMA).

L'idea di base del modello SARIMA (Seasonal ARIMA) è che il processo deve poter descrivere due tipi di relazioni all'interno della serie osservata; da una parte la correlazione tra valori consecutivi, ovvero quella modellata dagli usuali modelli ARIMA, e dall'altra la correlazione tra osservazioni che distano tra loro di un multiplo di periodo.

Per esplicitare il concetto ci si avvale della tabella 4.1 dove sono rappresentate le osservazioni relative a k anni di una serie storica mensile classificate in anno (in riga) e mese (in colonna).

Anno/mese	1	2	...	12
1	y_1	y_2	...	y_{12}
2	y_{13}	y_{14}	...	y_{24}
3	y_{25}	y_{26}	...	y_{36}
⋮	⋮	⋮	⋮	⋮
k	$y_{12(k-1)+1}$	$y_{11(k-1)+2}$...	y_{12k}

Tabella 4.1 Serie storica mensile classificata per anno e mese

La correlazione tra mesi successivi viene detta "orizzontale" mentre quella tra gli stessi mesi di anni diversi viene chiamata "verticale".

La correlazione orizzontale può venire modellata da un normale modello ARIMA (p, d, q) , ovvero:

$$\phi(B)(1-B)^d Y_t = \theta(B)b_t \quad (4.50)$$

Dove b_t non è un *white noise* dal momento che contiene ancora correlazione non spiegata, ovvero quella verticale che può essere modellata con il seguente modello ARIMA:

$$\Phi(B^S)(1 - B^S)^D b_t = \Theta(B^S)\varepsilon_t \quad (4.51)$$

Dove:

- $\Phi(B^S) = (1 - \phi_1 B^S - \phi_2 B^{2S} - \dots - \phi_p B^{pS})$ è un operatore autoregressivo di tipo stagionale
- $\Theta(B^S) = (1 - \theta_1 B^S - \theta_2 B^{2S} - \dots - \theta_q B^{qS})$ è il corrispondente operatore a media mobile stagionale
- S è il periodo stagionale, che nell'esempio è pari a 12
- $(1 - B^S)^D$ è l'analogo stagionale della differenziazione semplice
- $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ ovvero un *white noise*

Box e Jenkins, combinando insieme la (4.49) e la (4.50) hanno proposto di descrivere la struttura di dipendenza tra i periodi (ad esempio gli anni) e tra le stagioni (ad esempio i mesi) con il modello SARIMA $(p, d, q) \times (P, D, Q)_S$ così definito:

$$\phi(B)\Phi(B^S)(1 - B)^d(1 - B^S)^D Y_t = \phi_0 + \vartheta(B)\Theta(B^S)\varepsilon_t \quad (4.52)$$

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 ;
- $\vartheta(B)$ 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 ;
- $(1 - B^S)^D$ è l'operatore differenza stagionale di ordine D .

In modo analogo ai modelli ARIMA, se Y_t è generato da un SARIMA $(p, d, q) \times (P, D, Q)_S$ allora $Z_t = (1 - B)^d(1 - B^S)^D Y_t$ segue un modello SARMA $(p, 0, q) \times (P, 0, Q)_S$ che può essere scritto come un modello ARMA $(p + PS, q + QS)$. "Il modello SARMA $(p, 0, q) \times (P, 0, Q)_S$ è stazionario se le radici dell'equazione caratteristica $\phi(B)\Phi(B^S) = 0$ sono tutte, in modulo, maggiori di 1"; mentre "è invertibile se le radici dell'equazione $\vartheta(B)\Theta(B^S) = 0$ sono tutte esterne al cerchio unitario. Infine, comportamenti non stazionari, di tipo stagionale e/o stagionale, possono essere trattati ponendo, rispettivamente $d \neq 0$ e/o $D \neq 0$ ", anche se nella pratica si è soliti considerare $D \leq 1$. (Di Fonzo e Lisi, 2005).

La procedura di Box e Jenkins

Fino ad ora seguendo l'approccio moderno, si sono presentati i modelli solo dal punto di vista teorico. Tuttavia per costruire un modello che approssimi adeguatamente il processo generatore dei dati in esame, si è soliti seguire una

procedura proposta da Box e Jenkins (1976). Essa è composta dalle seguenti tre fasi (ad esclusione dei modelli SARIMA):

1. *Identificazione*: si specifica l'ordine del modello (ad esempio nell'ARIMA si specificano p,d,q) grazie alla funzione di autocorrelazione e la funzione di autocorrelazione parziale (entrambe stimate sul campione). L'idea di fondo è quella di riconoscere nella struttura della funzione di autocorrelazione empirica la struttura di una funzione di autocorrelazione teorica.

Tuttavia per evitare problemi di sovrapparametrizzazione (overfitting), è utile fare ricorso a criteri diversi per l'identificazione del modello che attribuiscono un costo all'introduzione di ogni nuovo parametro addizionale. Infatti è fondamentale sapere che un buon adattamento non conduce necessariamente ad una buona capacità previsiva del modello stesso. Come criterio si cita l'AIC (Asymptotic Information Criterion, Akaike, 1974) dato da:

$$AIC(K) = -\frac{2}{n}(\log L(\hat{\delta})) - k \quad (4.53)$$

dove k è il numero dei parametri del modello, $\hat{\delta}$ è il vettore (k x 1) che contiene i parametri stimati e $L(\hat{\delta})$ è la funzione di verosimiglianza (Azzalini, 1992) calcolata in $\hat{\delta}$ sotto l'ipotesi di gaussianità del white noise. La quantità $\frac{2k}{n}$ rappresenta una specie di penalità, legata al numero dei parametri del modello, se da un lato diminuisce al crescere di n, dall'altro diventa del tutto trascurabile per valori grandi di n. A causa di questo l'AIC tende a sovrapparametrizzare il modello. Per risolvere il problema si preferisce utilizzare un secondo criterio, detto SC (Schwarz Criterion, Schwarz, 1978):

$$SC(k) = \frac{2\log L(\hat{\delta})}{n} + k \frac{\log(n)}{n} \quad (4.54)$$

dove $k \frac{\log(n)}{n}$ è la penalità per l'inclusione di nuove variabili, che all'aumentare di n, decresce più lentamente di quello $\frac{2k}{n}$ usato da AIC. Per questa ragione SC porta all'identificazione di un modello più parsimonioso.

2. *Stima dei parametri*: i metodi più utilizzati sono il metodo dei *minimi quadrati non lineari* (Di Fonzo e Lisi, 2005) che non dipende da specifiche ipotesi distributive e il metodo della *massima verosimiglianza* (Hamilton, 1994) che richiede la conoscenza della distribuzione del termine di errore, ma fornisce stimatori con migliori

proprietà statistiche; anche se esistono molte altre metodologie di stima;

3. Controllo diagnostico: se il modello è stato correttamente identificato e stimato, allora sui residui $e_t = (y_t - \hat{y}_t)$ devono potersi riscontrare le ipotesi fatte a priori sui disturbi ε_t :
 - I. *Analisi grafiche*: se il modello prescelto è un'adeguata rappresentazione del processo generatore dei dati, la serie dei residui non dovrebbe mostrare alcun tipo di regolarità né valori particolarmente diversi gli uni dagli altri.
 - II. *Autocorrelazione dei residui*: la serie dei residui e_t può essere trattata come una nuova serie storica, per la quale può essere calcolata la funzione di autocorrelazione empirica. "Per ogni valore di k è possibile utilizzare $\hat{\rho}_k$ per verificare se ρ_k è significativamente diversa dalla autocorrelazione di un processo *white noise* che è nulla. Poiché per un *white noise* la varianza di $\hat{\rho}_k$ è approssimativamente n^{-1} per ogni k , per ogni ipotesi di normalità la regione di accettazione al livello di significatività del 5% è data da $[-\frac{1.96}{\sqrt{n}}, +\frac{1.96}{\sqrt{n}}]$. Un valore di $\hat{\rho}_k$ fuori da questo intervallo porta a ritenere ρ_k significativamente diverso da 0" (Di Fonzo e Lisi, 2005)
 - III. *Test di tipo portmanteau*: per una verifica complessiva di assenza di correlazione (si veda Di Fonzo e Lisi, 2005)
 - IV. *Verifica di casualità dei residui* (si veda Siegel e Castellan, 1992).
 - V. *Test di normalità dei residui*: verifica se i residui sono distribuiti normalmente, poiché se così fosse l'incorrelazione dei residui implica anche la loro indipendenza (vedasi Di Fonzo e Lisi, 2005).

Va sottolineato il fatto che per generici modelli SARIMA $(p, d, q) \times (P, D, Q)_S$ l'identificazione del modello a partire dalle funzioni di autocorrelazioni richiede lo stesso tipo di procedura proposta da Box e Jenkins (1976) descritta precedentemente, ma con alcuni accorgimenti. Infatti dal momento che riconoscere il modello in un unico passo a partire dalle funzioni di autocorrelazione risulta spesso difficoltoso, si è soliti procedere a più passi. Si identifica quindi la componente stagionale e poi quella non stagionale (o viceversa a seconda dei casi). Anche in questo caso vale quanto detto per i criteri AIC o SC, ovvero per l'identificazione dei modelli SARIMA si sceglie il modello che minimizza tali criteri.

Utilizzo predittivo dei modelli

Dopo aver identificato e stimato, mediante la procedura di Box e Jenkins, un modello usufruendo dell'informazione fino all'istante $T = n$, lo si vuole utilizzare per prevedere il valore della serie al tempo $n + k$. La previsione \widehat{y}_{n+k}

di y_{n+k} è il particolare valore che assume il previsore \widehat{Y}_{n+k} all'istante $n+k$ ed è data da:

$$\widehat{y}_{n+k} = E[Y_{n+k} | Y_n = y_n, Y_{n-1} = y_{n-1}, \dots, Y_1 = y_1, Y_0 = y_0, \dots] \equiv E_n[Y_{n+k}] \quad (4.55)$$

ovvero dal valore atteso di Y_{n+k} condizionato all'informazione passata.

Teoricamente, il condizionamento nella (4.55) dovrebbe riguardare tutte le infinite v.c. Y_t che compongono il processo fino all'istante n . Ma, nella pratica, ciò non è quasi mai possibile e quindi il condizionamento avviene solamente rispetto all'informazione realizzata nel passato. Si considera quindi l'approssimazione:

$$E[Y_{n+k} | Y_n = y_n, \dots, Y_1 = y_1, Y_0 = y_0, Y_{-1} = y_{-1}, \dots] \approx E[Y_{n+k} | Y_n = y_n, \dots, Y_1 = y_1] \quad (4.56)$$

Nel seguito non verranno distinti i due casi e si parlerà semplicemente di valore atteso condizionato.

In generale vale la seguente regola:

$$E[Y_{n+j} | Y_n = y_n, \dots, Y_1 = y_1] = \begin{cases} y_{t+j} & j \leq 0 \\ \hat{y}_{t+j} & j > 0 \end{cases} \quad (4.57)$$

E prendendo in considerazione l'errore di previsione $e_t = y_t - \hat{y}_t$,

$$E[\varepsilon_{t+j} | Y_n = y_n, \dots, Y_1 = y_1] = \begin{cases} e_{t+j} & j \leq 0 \\ 0 & j > 0 \end{cases} \quad (4.58)$$

In termini generali, l'errore che si commette utilizzando \widehat{y}_{t+k} al posto di y_{t+k} considerando solo l'informazione disponibile fino all'istante t , viene indicato con:

$$e_{t+k} = y_{t+k} - \widehat{y}_{t+k} \quad (4.59)$$

Poiché e_{t+k} è una v.c. è possibile studiarne la struttura probabilistica al fine di valutare la precisione della previsione, per esempio utilizzando una delle misure di accuratezza previsionale presentate nel capitolo 3.

4.5 Modelli di regressione

Un generico modello di regressione può essere definito nel seguente modo:

$$y = g(x_1, \dots, x_p; \varepsilon) \quad (4.60)$$

dove y è la variabile risposta che può essere quantitativa (continua o discreta) o qualitativa (ordinale, nominale, dicotomica o politomica), mentre le x_1, \dots, x_p sono le variabili esplicative di natura quantitativa, qualitativa o mista.

La variabile risposta è l'effetto congiunto di due componenti:

- *sistemica*: esprime la relazione della variabile risposta con le variabili esplicative,
- *casuale* (o accidentale): spiega la variabilità di y non spiegata dalle variabili esplicative.

Per effetto del suo legame con ε , la variabile risposta è una variabile casuale con una distribuzione di probabilità f che dipende dalle variabili esplicative:

$$Y \sim f(y; x_1, \dots, x_p) \quad (4.61)$$

A seconda della natura delle variabili coinvolte, possono essere costruiti diversi modelli rappresentati nella tabella 4.2.

Variabile risposta (Y)		Variabili esplicative (x_1, \dots, x_p)		
		quantitative	nominali	miste
quantitativa	continua	Modello lineare (reg. multipla)	Modello lineare (ANOVA)	Modello lineare (ANCOVA)
	discreta	reg. di Poisson	reg. di Poisson; modello log-lineare	reg. di Poisson
nominale	politomica	regressione multinomiale	regressione multinomiale; modello log-lineare	regressione multinomiale
	dicotomica	regressione binomiale (es. logistica)	regressione binomiale (es. logistica); modello log lineare	regressione binomiale (es. logistica)

Tabella 4.2 Classificazione di alcuni modelli statistici in base alla natura delle variabili coinvolte

Per la definizione di un modello di regressione, si assume di osservare un campione di T istanti temporali, dove il generico istante t è descritto come:

$$(x_{t1}, \dots, x_{tp}, y_t) \quad t = 1, \dots, T$$

Le quantità osservate saranno, pertanto:

- una matrice non stocastica $X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{T1} & \cdots & x_{Tp} \end{pmatrix}$
- un vettore $y = (y_1, \dots, y_t, \dots, y_T)^T$ che si suppone essere la realizzazione di un vettore aleatorio Y , ovvero di una casuale multipla:

$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_T \end{pmatrix}$ dove ciascuna componente Y_t ($t = 1, \dots, T$) è essa stessa una variabile casuale.

4.5.1 Modello di regressione lineare

Il modello di *regressione lineare* è uno dei modelli più semplici per la previsione della domanda, che consiste nello studio delle relazione tra p variabili esplicative e una variabile risposta y , che è appunto la domanda stessa.

A partire dalla (4.60) si definisce il modello di regressione lineare nel seguente modo:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon \quad (4.62)$$

dove la variabile risposta y è una variabile quantitativa; mentre le variabili esplicative x_j ($j = 1, \dots, p$) possono essere quantitative o categoriali (fattori).

Va sottolineato come rispetto al caso generale la componente casuale entra in modo additivo nel modello; la componente sistematica è una funzione lineare dei parametri β_1, \dots, β_p .

Tuttavia la (4.62) non impone che la relazione sia lineare rispetto alle variabili, ma soltanto rispetto ai parametri.

Date T osservazioni $(x_{t,1}, \dots, x_{t,p}, y_t)$ per $t = 1, \dots, T$; si assume che ciascuna y_t sia una realizzazione casuale Y_t , rappresentabile attraverso la seguente relazione:

$$Y_t = \beta_1 x_{t1} + \dots + \beta_p x_{tp} + \varepsilon_t \quad \text{per } t = 1, 2, \dots, T \quad (4.63)$$

Una forma di scrittura più compatta e semplice della precedente è data dalla notazione matriciale:

$$Y = X\beta + \varepsilon \quad (4.64)$$

Che in forma estesa risulta essere:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_T \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{T1} & \dots & x_{Tp} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{pmatrix} \quad (4.65)$$

dove:

- Y è un vettore di dimensione $(T \times 1)$ che contiene i valori della serie di domanda D_t^B riscontrati lungo tutto l'orizzonte temporale, rispettivamente per i periodi $t = 1, \dots, T$
- X è una matrice di dimensione $(T \times p)$, detta matrice di regressione contenente tutte le variabili esplicative
- β è il vettore $(p \times 1)$ dei parametri (coefficienti) di regressione

- ε è un vettore casuale ($T \times 1$)

Le ipotesi del secondo ordine per la (4.64) suggeriscono che:

- X è una matrice non stocastica di rango pieno $p \leq T$;
- $E(\varepsilon) = 0$, ovvero la componente di errore non è attribuibile a fattori sistematici, quindi è ragionevole assumere che abbia media nulla;
- $VAR(\varepsilon_t) = \sigma^2 I_T$, ovvero gli errori sono incorrelati.

Dalle ipotesi del secondo ordine segue che i primi 2 momenti sono:

- $E(Y) = E(X\beta + \varepsilon) = X\beta = \mu$ (4.66)

- $VAR(Y) = VAR(X\beta + \varepsilon) = VAR(\varepsilon) = \sigma^2 I_T$ (4.67)

Va sottolineato che se il modello lineare presenta l'intercetta, allora la variabile esplicativa associata viene descritta da un vettore $(1, \dots, 1)^T$, mentre la (4.65) diventa:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_T \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{T1} & \cdots & x_{Tp} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{pmatrix} \quad (4.68)$$

dove questa volta X e β hanno rispettivamente dimensioni $(T \times (p + 1))$ e $((p + 1) \times 1)$.

Per ottenere uno stimatore del parametro β senza ricorrere ad assunzioni sulla distribuzione di probabilità di Y ¹¹ una metodologia molto comune in letteratura è l'utilizzo del *metodo dei minimi quadrati*. Questo porta ad individuare quei valori per cui si consegue il minimo rispetto a β della seguente funzione obiettivo:

$$D(\beta) = \|y - \mu\|^2 = (y - X\beta)^T (y - X\beta) = \sum_{t=1}^T (y_t - (\beta_1 x_{t1} + \cdots + \beta_p x_{tp}))^2 \quad (4.69)$$

dove $\| \cdot \|$ indica la norma euclidea del vettore, cioè la radice della somma dei quadrati degli elementi.

Il risultato di questo problema di minimo è $\hat{\beta} = (X^T X)^{-1} X^T y$ in corrispondenza del quale si individuano i valori interpolati:

$$\begin{aligned} \widehat{y}_t &= X\hat{\beta} \\ \widehat{y}_t &= \widehat{\beta}_0 + \widehat{\beta}_1 x_{t1} + \cdots + \widehat{\beta}_p x_{tp} \end{aligned} \quad \text{per } t = 1, \dots, T \quad (4.70)$$

Ne consegue che la previsione della domanda \widehat{y}_{t+k} per i periodi futuri dell'intervallo di previsione $t + k$, $k = 1, 2, 3, \dots$, si ottiene come:

¹¹ Se oltre alle ipotesi del secondo ordine si aggiunge l'ipotesi $\varepsilon = N_n(0, \sigma^2 I_n)$, considerando la (4.63), segue che la distribuzione di probabilità è $Y \sim N_n(X\beta, \sigma^2 I_n)$; si parla quindi di ipotesi di normalità del modello lineare grazie alle quali si possono costruire procedure inferenziali esatte

$$\widehat{y_{t+k}} = \widehat{\beta_0} + \widehat{\beta_1}x_{t+k,1} + \dots + \widehat{\beta_p}x_{t+k,p} \quad (4.71)$$

Una volta stimato il modello lineare, gli strumenti diagnostici vengono utilizzati per individuare eventuali indizi di scostamento dalle ipotesi alla base dello stesso. Alcuni di questi sono:

- *coefficiente di determinazione*: per valutare il grado di adattamento del modello definito come: $R^2 = 1 - \frac{\sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - \bar{y})^2}$ dove $\bar{y} = \sum_t \frac{y_t}{T}$ indica la media aritmetica delle y_t ;
- grafico dei residui $\hat{\varepsilon}^{12}$ contro i valori previsti $\hat{\mu}$, oppure contro le variabili esplicative: si verifica che il comportamento dei residui non si discosti dal comportamento che avrebbero se le ipotesi alla base del modello fossero soddisfatte (andamento regolare intorno allo 0);
- grafici di y contro ciascuna variabile esplicativa: aiuta a capire come ciascuna delle variabili esplorative deve entrare nel modello;
- grafico della variabile aggiunta: utile a isolare l'effetto di ciascuna variabile esplicativa su y , in particolare se correlate.

Qualora le ipotesi del modello di regressione lineare non fossero soddisfatte, e non si ritenga accettabile uno scostamento da tali ipotesi, è necessario rimediare modificando il metodo di stima, adattando il modello (cioè con modifiche minori dello stesso) o utilizzando un modello diverso, capace di conformarsi ai dati a disposizione.

In particolare nel caso di violazione dell'ipotesi di linearità alcuni rimedi possono essere:

- trasformazioni delle variabili coinvolte;
- modelli lineari generalizzati (GLM);
- modello non lineare.

4.5.2 Modello di regressione lineare con trasformata logaritmica

Nel caso in cui il modello è non lineare, e quindi viola l'ipotesi di linearità del modello di regressione lineare, la ricerca di trasformazioni per linearizzare la relazione può seguire due prospettive. La prima consiste nel basarsi sulla teoria sottostante al fenomeno che Y e X descrivono (ad esempio una teoria economica), la seconda si basa sul fatto di studiare la relazione tra le quantità sulla base dell'analisi esplorativa iniziale o dell'analisi dei residui.

Tenuto conto che l'ipotesi di linearità si riferisce ai parametri e non alle variabili, sono ammissibili situazioni in cui $E(Y)$ è legata linearmente a una o più trasformate di X . In formule risulta:

¹² I residui sono il surrogato degli errori ε_t , non osservabili, e si definiscono come: $\hat{\varepsilon}_t = y_t - \hat{y}_t$ per $t = 1, \dots, T$

$$Y_t = \beta_1 + \beta_2 g_2(\tilde{x}_t) + \dots + \beta_p g_p(\tilde{x}_t) + \varepsilon_t \quad (4.72)$$

dove g_2, \dots, g_p sono funzioni note.

Bisogna evidenziare il fatto che l'utilizzo di trasformazioni non lineari di variabili, non ha effetti sull'impianto teorico del modello lineare descritto in precedenza, salvo il caso in cui si operino trasformazioni della variabile risposta. La funzione obiettivo (4.69) e, di conseguenza il criterio di ottimalità perseguito, opera sulla scala trasformata.

Una trasformazione frequentemente usata è quella logaritmica. In molti casi è sensato trasformare oltre alle variabili esplicative, anche la variabile risposta. Questo tipo di trasformazione viene spesso utilizzata quando si lavora con quantità intrinsecamente positive come la domanda commerciale, il prezzo o lo sconto. "Questa trasformazione ha il vantaggio di operare con variabili che variano in $(-\infty, +\infty)$, cioè il giusto supporto per i modelli lineari. Questo fatto a sua volta comporta che, una volta invertita la funzione, siamo sicuri di ottenere quantità positive per i valori previsti della variabile risposta. Infine un ulteriore vantaggio della trasformazione logaritmica è che spesso corregge l'eteroschedasticità dei residui" (Azzalini e Scarpa, 2004).

Ad esempio considerando il modello di partenza:

$$Y_t = \beta_0 x_{t2}^{\beta_2} e^{\beta_3 x_{t3}} \varepsilon_t \quad \text{per } t = 1, \dots, T$$

esso è linearizzabile grazie alla trasformazione di alcune variabili esplicative e della variabile risposta. In formule risulta:

$$\log(Y_t) = \log(\beta_0) + \beta_2 \log(x_{t2}) + \beta_3 x_{t3} \quad (4.73)$$

In questo specifico esempio, la previsione della domanda \widehat{y}_{t+k} per i periodi futuri dell'intervallo di previsione $t+k$, $k = 1, 2, 3, \dots$, si ottiene come:

$$\widehat{y}_{t+k} = \widehat{\beta}_0 + \widehat{\beta}_1 \log(x_{t+k,1}) + \dots + \widehat{\beta}_p x_{t+k,p} \quad (4.74)$$

4.6 Modelli di previsione avanzati: Rete Neurale

Il termine Rete Neurale si riferisce a un'ampia famiglia di tecniche particolarmente sviluppate nel contesto del *machine learning*.

Generalmente questo metodo necessita di molte più osservazioni rispetto ai metodi descritti fino ad ora, ma permette una maggiore flessibilità.

Nella sua versione base, più semplicistica, la Rete Neurale mette in relazione delle variabili esplicative di input p con q variabili di risposta o di output.

Tuttavia queste variabili vengono filtrate grazie alla presenza di uno strato di r variabili latenti, cioè non osservate. Il procedimento logico è il seguente:

- le variabili di input influenzano le variabili latenti;
- le variabili latenti influenzano a loro volta le variabili di output.

Va evidenziato che mentre le variabili di input e output, sono determinate per definizione dal problema della previsione della domanda, le variabili dello strato latente devono essere imposte dai Demand Planner.

In generale una rete neurale è definita come uno schema di regressione a due stadi, generalmente di tipo non lineare o almeno parzialmente non lineare. Si indicano con x_h, z_j e y_k rispettivamente la generica variabile di input, latente e di output, a cui vanno aggiunte le variabili x_0 e z_0 , che risultano identicamente pari a 1 (Azzalini e Scarpa, 2004: pag.101).

La rete neurale si può quindi esprimere attraverso le seguenti due equazioni:

$$z_j = f_0\left(\sum_{h \rightarrow j} \alpha_{hj} * x_h\right) \quad (4.75)$$

$$y_k = f_1\left(\sum_{j \rightarrow k} \beta_{jk} * z_j\right) \quad (4.76)$$

dove α_{hj} e β_{jk} sono parametri da stimare.

Per determinare f_0 e f_1 , chiamate funzioni di attivazione, si è soliti utilizzare rispettivamente una funzione logistica e una relazione di tipo lineare. In formule risulta:

$$f_0(\mu) = \frac{e^\mu}{1+e^\mu} \quad (4.77)$$

$$f_1(u) = u \quad (4.78)$$

Va evidenziato il fatto che almeno una delle due funzioni deve essere di tipo non lineare, questo serve ad evitare che la rete si riduca ad un insieme di relazioni lineari eliminando di fatto lo strato latente.

Restano da determinare il numero r di unità nello strato latente, e l'insieme dei coefficienti α e β presenti nella (4.75) e (4.76). Purtroppo non esistono criteri facilmente utilizzabili per la scelta del numero r , se non quello di provarne alcuni confrontando i risultati ottenuti. Per questa ragione si assume r come fissato per il calcolo dei coefficienti α e β , sulla base di osservazioni campionarie, che si realizza rendendo minima la seguente funzione obiettivo:

$$D = \sum_i || y^{(i)} - f(x^{(i)}) ||^2 \quad (4.79)$$

dove $y^{(i)}$ indica il vettore q -dimensionale di variabili risposta relative alla i -ma osservazione, e analogamente $x^{(i)}$ è il corrispondente vettore p -dimensionale della variabili esplicative, mentre $f(x)$ è il vettore la cui k -ma componente è:

$$f(x)_k = f_1\left\{ \sum_{j \rightarrow k} \beta_{jk} f_0\left(\sum_{h \rightarrow j} \alpha_{hj} x_h\right) \right\} \quad \text{per } k = 1, \dots, q \quad (4.80)$$

Per evitare problemi di sovra-adattamento, si applica un termine di penalizzazione alla funzione obiettivo definita nella (4.78) ottenendo funzioni del tipo:

$$D_0 = D + \lambda J(\alpha, \beta) \quad (4.81)$$

dove λ è un parametro positivo di regolazione che si è soliti porre pari ad un valore compreso nell'intervallo 10^{-4} e 10^{-2} seguendo le indicazioni di Venables e Ripley (1997); mentre $J(\alpha, \beta)$ è una funzione di penalizzazione.

Una tra le forme più comuni, presenti in letteratura, di quest'ultima funzione è chiamata *weight decay* ed è caratterizzata dal fatto di tendere a stringere i parametri verso lo 0. In formule è definita come:

$$J(\alpha, \beta) = \|\alpha\|^2 + \|\beta\|^2 \quad (4.82)$$

Il metodo più comunemente utilizzato per minimizzare D_0 è un procedimento di ottimizzazione numerica chiamato *back-propagation*, che gode di interessanti proprietà.

Tuttavia bisogna sottolineare il fatto che la funzione obiettivo D_0 ha spesso molti punti di minimo locale, e per questo è opportuno effettuare molteplici tentativi nell'avviare l'algoritmo di ottimizzazione numerica da diversi punti iniziali.

4.7 Modelli per il new product forecasting

Dal punto di vista della generazione della previsione commerciale, risulta molto complicato utilizzare i modelli di sales forecasting descritti in precedenza per generare il sales forecast per i prodotti che, per definizione, non possiedono una serie storica.

Nella GDO esistono due tipologie in base alle quali suddividerli:

- prodotti per i quali non esistono prodotti simili, perché non sono mai stati prodotti nel mondo o perché non è mai stato costituito un accordo commerciale con il fornitore di quella particolare tipologia di prodotto;
- prodotti che portano piccoli cambiamenti rispetto a prodotti già esistenti, ad esempio l'introduzione di un nuovo gusto di yogurt oppure di una nuova fragranza di shampoo.

Silver et al. (1998) sostengono che quando le serie storiche contengono poche o nessuna osservazione, un forecast di tipo qualitativo basato solo sul giudizio e l'esperienza dei Demand Planner pare abbia senso.

Tuttavia per prevedere nuovi prodotti può essere utilizzato un processo strutturato che combina analisi di similitudini, time series analysis e conoscenza dei Demand Planner.

Il processo di previsione di un nuovo prodotto segue le seguenti fasi (vedasi Chase, 2013):

1. *Query*. Si cercano dei prodotti che abbiano delle caratteristiche simili al nuovo prodotto, e si considerano le serie storiche di questi prodotti che risulteranno essere simili alla serie storica del nuovo prodotto che non è conosciuta;
2. *Filtraggio*. Rimozione dei prodotti meno simili al nuovo prodotto dalla lista restituita nella fase 1.
3. *Modellazione*. Estrazione delle caratteristiche dei modelli statistici dalla lista delle serie storiche output della fase 2.
4. *Previsione*. Uso delle caratteristiche dei modelli statistici per prevedere i nuovi prodotti

Fase 1. Lo scopo di questa fase è quello di trovare una lista di prodotti, con caratteristiche simili al nuovo prodotto, che possano essere utili nel prevedere la serie storica sconosciuta del nuovo prodotto. Infatti si presume che le serie scelte siano simili a quella del nuovo prodotto. Tuttavia poiché la serie storica del nuovo prodotto non è conosciuta, vengono utilizzate solamente le caratteristiche conosciute del nuovo prodotto che i Demand Planner selezionano in base alla loro esperienza. Queste possono essere, ad esempio, l'imballo, la categoria di appartenenza, il brand, il prezzo.

Tipicamente la lista dei prodotti simili non deve essere né troppo piccola né troppo grande poiché il campione potrebbe non essere rappresentativo delle proprietà e/o delle variazioni associate alla serie del nuovo prodotto.

Fase 2. Lo scopo di questa fase è quello di rimuovere quei prodotti che sono meno simili rispetto al nuovo prodotto dalla lista restituita nella fase 1.

I filtri possono essere implementati in vari modi:

- **dimension reduction analysis** che riduce ciascuna serie storica di output della fase 1 a un insieme di caratteristiche che non dipendono dal tempo;
- **similarity analysis** che misura la similarità tra due serie rispetto al tempo.

Altre tecniche statistiche sono **misura delle distanza, metodi di clustering** o semplicemente **regole di business**.

Per esempio può essere specificato un modello di decadimento esponenziale e poi clusterizzare i parametri di tale modello.

I Demand Planner scelgono quindi qual è il cluster più rappresentativo delle caratteristiche della serie storica del nuovo prodotto, eliminando eventuali outliers.

Fase 3. Le serie risultanti dalla fase 2, essendo continue e regolari vengono modellate usando una tecnica di lisciamento di gruppo detta “pooled smoothing”, utilizzando i metodi visti nel capitolo 3 per la suddivisione delle serie storica di domanda con cui calcolare la previsione (la serie viene stimata sulla parte di training set e valutata sul test set).

I Demand Planner quindi valutano il modello attraverso una delle misure di accuratezza previsionale descritte nel capitolo 3 e, se lo ritengono abbastanza soddisfacente, passano alla fase 4. Altrimenti decidono di levare una delle serie e di iniziare nuovamente la fase 3. Anche se “non ci si dovrebbe aspettare una maggiore accuratezza per il semplice fatto di avere eliminato una serie con un errore di previsione elevato” (Chase, 2013); ma talvolta funziona se la conoscenza dei Demand Planner è particolarmente elevata.

Fase 4. Lo scopo di questa fase è quello di prevedere nuovi prodotti utilizzando le previsioni pooled (output della fase 3) associate con le caratteristiche individuate dall’insieme delle serie storiche output della fase 2.

4.8 Combinazione di modelli di Sales Forecasting

Non sempre le diverse metriche di errore sono concordi nel definire il miglior modello predittivo (internamente ottimizzato dal best fit-parametrico) per un determinato SKU (vedasi capitolo 3).

Per migliorare il best-fit algoritmico si è soliti generare una previsione combinata del forecast dei diversi modelli di sales forecasting, infatti “studi sperimentali condotti su numerose serie storiche continue hanno dimostrato l’efficacia di tale approccio, al fine di incrementare l’accuratezza complessiva erogata dal sistema di sales forecasting” (Milanato, 2008).

Ciascun algoritmo di sales forecasting z_j genera una previsione $\hat{y}_t(z_j)$ per il generico futuro periodo $t > T$, per tutti i periodi dell’intervallo di previsione. La previsione combinata CF_t (detta *combined sales forecast*) è essenzialmente una combinazione lineare pesata delle previsioni di domanda generate dai singoli modelli di *sales forecasting*. In formule:

$$CF_t = \frac{\sum_{j \in Z} w(z_j) * \hat{y}_t(z_j)}{\sum_{j \in Z} w(z_j)} \quad \forall t > T \quad (4.83)$$

dove i pesi attribuiti alla previsione dai singoli modelli predittivi sono $w(z_j)$.

La previsione viene quindi ottenuta come media pesata, normalizzata al valore unitario, dei contributi offerti dai singoli modelli di sales forecasting.

Affinché questa metodologia abbia successo occorre quindi definire i pesi $w(z_j)$ da assegnare a ciascun algoritmo usato nella combinazione.

I più utilizzati sono:

- **pesi uniformi:** garantiscono una sufficiente accuratezza in quanto compensano errori elevati di alcuni modelli. In formule si ottiene il seguente *combined sales forecast*:

$$CF_t = \frac{\sum_{j \in Z} \widehat{y}_t(z_j)}{\text{car}\{Z\}} \quad (4.84)$$

- **pesi inversamente proporzionali** ad e_t , calcolato nel periodo di test relativo (tale per cui $T = t$), commesso da ciascun modello nell'eseguire la previsione all'istante t per i periodi passati $t, t - 1, t - 2, \dots$

Detto in altri termini, "ciascun modello assume, nella previsione futura eseguita a partire dal primo periodo futuro $t + 1$, un peso maggiore quanto migliore si è dimostrata la sua accuratezza predittiva nel calcolo del sales forecast nell'intervallo passato di test" (Milanato, 2008: pag. 434). Quindi, per esempio, utilizzando la metrica dell' RMSE¹³ per la valutazione dell'accuratezza¹⁴ di ciascun algoritmo z_j , la formula di calcolo della previsione combinata diventa:

$$CF_t = \frac{\sum_{j \in Z} \frac{\widehat{y}_t(z_j)}{RMSE(z_j)}}{\sum_{j \in Z} \frac{1}{RMSE(z_j)}} \quad \forall t > T \quad (4.85)$$

- **pesi inversamente proporzionali** ad e_t , ovvero all'errore di previsione commesso da ciascun modello utilizzato nell'eseguire la previsione all'istante t sull'ultimo periodo storico $T = t$.

"In questo caso, viene misurato l'errore di previsione solamente sull'ultimo dato disponibile di domanda, in modo incrementale rispetto all'ultima edizione del sales forecast" (Milanato, 2008). In formule la previsione combinata risulta essere:

$$CF_t = \frac{\sum_{j \in Z} \frac{\widehat{y}_t(z_j)}{RMSE_T(z_j)}}{\sum_{j \in Z} \frac{1}{RMSE_T(z_j)}} \quad \forall t > T \quad (4.86)$$

¹³ Vedasi capitolo 3 per la sua definizione

¹⁴ In questo caso solamente misure assolute di *forecast error* possono essere utilizzate per la valutazione dell'accuratezza

5. CASO DI STUDIO

Come si è visto nei precedenti capitoli l'obiettivo principe dei Demand Planner è quello di prevedere la domanda commerciale sia di tipo regolare che promozionale.

In questo elaborato si è posta l'attenzione più sul secondo tipo di problema in quanto meno sviluppato nella letteratura e quindi più problematico da trattare.

Per il caso di studio si sono utilizzati dati reali provenienti da una importante azienda del settore GDO leader nel Triveneto.

Questa da molti anni ha deciso di puntare su campagne promozionali, utili al duplice scopo di attrarre nuovi clienti ma anche di mantenere i propri, piuttosto che puntare sulla varietà di assortimento come scelto invece da alcune aziende competitor.

Tra le varie tipologie di promozioni ideate dell'azienda si elencano:

- **Volantini:** vengono rappresentati graficamente i prodotti che mirano ad attrarre maggiormente i clienti in quello specifico momento. Sono prodotti, scelti dopo diverse analisi dei volantini della concorrenza, che risultano essere più convenienti e con la presenza di marchi Leader di mercato che hanno l'obiettivo di aumentare il fatturato aziendale;
- **Offerte interne:** sono prodotti che non vengono rappresentati graficamente da un volantino, ma vengono comunicati solo nei punti vendita. Questi prodotti creano il mix merceologico per coprire i fabbisogni del cliente. Sono opportunità di acquisto che il fornitore dà all'azienda e vengono proposte come opportunità al cliente finale;
- **Cataloghi/Inseriti:** sono prodotti che rappresentano un particolare tema che può essere legato alla stagionalità dei prodotti o ad una festa particolare o ad un evento. Ad esempio Catalogo Mobile-Giardino o Catalogo Caldo Casa o Festa del Tonno. La differenza tra catalogo e inserto risiede solamente nella rappresentazione grafica che nel primo caso consiste in un volantino specifico e che serve per comunicare il vasto assortimento per quel particolare tema, mentre il secondo, l'inserto, è costituito da pagine inserite all'interno di un volantino bisettimanale e che richiamano l'attenzione su un tema senza però comunicare la profondità assortimentale;
- **Sponsor:** sono prodotti che possono a volte essere rappresentati graficamente e che aiutano l'azienda a fidelizzare il cliente tramite delle campagne di raccolta punti con erogazione di un premio.

Per semplicità di trattazione si prenderanno in considerazione solo SKU appartenenti alla promozione "volantino".

Dati alla mano questo tipo di promozione risulta particolarmente redditizia per l'azienda. Infatti, per esempio, per uno stesso SKU lo scostamento tra venduto promozionale e regolare risulta molto più marcato rispetto agli altri tipi di promozioni.

Tuttavia va sottolineato il fatto che quanto verrà discusso può essere implementato in analogia anche per gli SKU che appartengono agli altri tipi di promozione.

Oltre alle promozioni già citate presenti nell'azienda, il mondo promozionale è in continua evoluzione e nuove tipologie di offerte vengono ideate continuamente dalla Marketing Intelligence.

A causa di ciò nasce un problema molto comune e difficile da risolvere per i Demand Planner, ovvero la **cannibalizzazione**. Questo accade quando due o più SKU, che appartengono a promozioni differenti o semplicemente classi diverse (classe promozionale o classe regolare), sono ritenuti simili per un generico consumatore. Quest'ultimo tendenzialmente acquisterà lo SKU, tra quelli che ritiene sostitutivi, che massimizzerà la propria funzione di utilità che di solito coincide con la scelta dello SKU con prezzo inferiore (anche se non è detto che il consumatore segua sempre questa logica).

5.1 Fasi del processo:

1. Input:

- *File Gestionale*: Informazioni di carattere anagrafico con dettaglio SKU-PDV (Punto di Vendita) relative alla specifica promozione per cui si richiede di fare la previsione (sell-in e sell-out ben definiti, per il volantino sono 2 settimane). Queste informazioni sono raccolte in un file gestito in autonomia da Marketing Intelligence e Ufficio Acquisti. La prima area prende decisioni in merito alla numerica e all'area merceologica degli SKU in volantino tenendo conto dello spazio disponibile sul supporto cartaceo, per poi stamparlo; mentre la seconda sui singoli SKU da inserire in offerta con il relativo sconto percentuale.
- *File DWH*: Informazioni relative alla serie storiche settimanali del venduto con dettaglio SKU-PDV più informazioni di carattere anagrafico delle settimane precedenti. Queste informazioni sono contenute nel DWH aziendale e vengono estratte attraverso specifiche query implementate con linguaggio SQL.

2. **R:** strumento open source che grazie a degli script sviluppati e implementati ad hoc dai Demand Planner permette di produrre il forecast richiesto, utilizzando tecniche di modellazione statistica.
3. **Output:** Previsioni aggregate per SKU-totale PDV ma poi spaccate per singole combinazioni di SKU-PDV, utilizzando regole di business e scelte manageriali.
4. **Prenotazione:** fase in cui ogni PDV può prenotare le quantità dei singoli SKU. Per prenotare questi quantitativi i PDV dovrebbero seguire il seguente ragionamento: se la stima della quantità di merce che avrà in casa al momento del sell-in della promozione (giacenza) è minore di quella che presume di vendere durante la promozione, allora ordinerà il quantitativo rappresentante quella differenza; in caso contrario non dovrebbe procedere alla prenotazione.
5. **Riordino:** fase in cui sulla base delle singole prenotazioni vengono prese decisioni sul rifornimento dei magazzini principali dell'azienda da cui si riforniranno i singoli PDV. In particolare si decide il quantitativo complessivo di merce da comprare per ogni SKU in modo da coprire il periodo promozionale, evitando rotture di stock a magazzino, ma allo stesso tempo evitando di intasarlo (overstock). Ovviamente per il rifornimento del magazzino centrale vanno considerate le prenotazioni non soltanto a livello aggregato di PDV-promozione, quindi non solo quelle legate ad una specifica offerta, ma anche le prenotazioni di periodi regolari. L'obiettivo è quindi ridurre al minimo il trade off: rotture di stock contro l'overstock di magazzino.
6. **Evasione:** il quantitativo ordinato dai PDV viene loro consegnato secondo regole di natura logistica e scelte manageriali per coprire la promozione. Ad esempio, non viene consegnata tutta la merce prenotata dai PDV prima che inizi la promozione, ma viene divisa in alcune trance, per evitare l'intasamento del magazzino del singolo PDV.
7. **Vendita:** fase coincidente rispettivamente con inizio nel sell-in dell'offerta e fine nel sell-out dell'offerta.

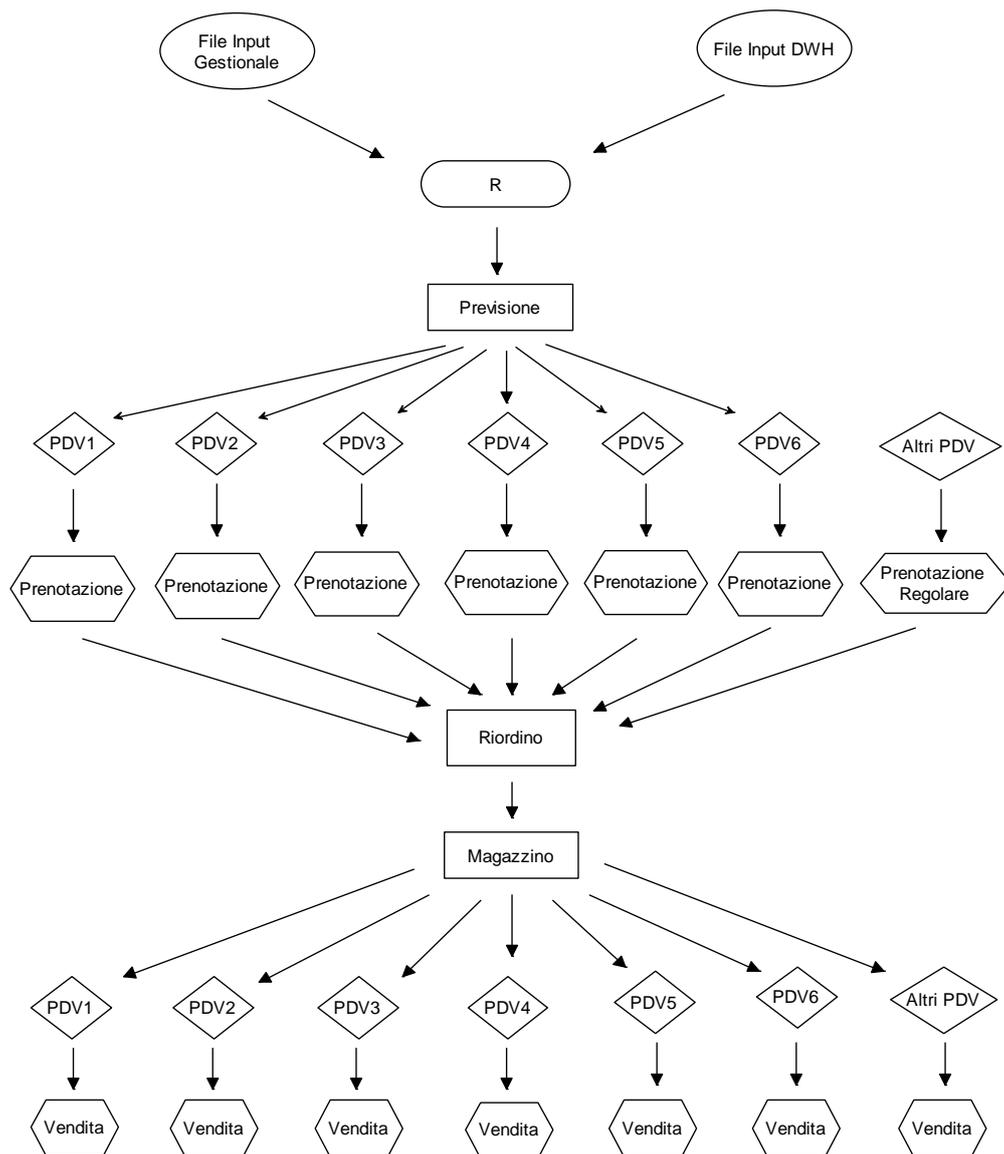


Figura 5.1

Chiarito il processo aziendale, mostrato in figura 5.1, si procede descrivendo la fase di produzione della previsione, ovvero la metodologia implementata per rispondere alla richiesta di prevedere la domanda commerciale promozionale che è l'obiettivo dei Demand Planner, oltre che un focus sulle problematiche relative all'accuratezza previsionale nel contesto aziendale.

Per fare ciò verranno descritte in generale le logiche principali utilizzate a livello del singolo SKU.

Successivamente verranno considerati casi specifici di singoli SKU presi come campione rappresentativo da una promozione di volantino.

5.2 Analisi preliminari e codifica variabili

Utilizzando il software open source R e in particolare la libreria "ROracle" si è creata una connessione con il DWH in modo da interrogarlo tramite query implementate con il linguaggio html, per ottenere il file di input DWH.

L'estrazione coinvolge il venduto settimanale dall'inizio del 2011 al periodo corrente (fine 2017) per ogni SKU per cui è richiesta una previsione. La lista degli SKU si trova nel file di input Gestionale dove, come già spiegato, gli SKU vengono scelti in sinergia tra Marketing Intelligence (che decide il numero e la tipologia di prodotti) e l'ufficio Acquisti (che decide il singolo prodotto da inserire rispettando le direttive del Marketing).

In particolare il file Gestionale di input della promozione volantino presa in esame ha un ordine di grandezza pari a circa $1.4 \cdot 10^4$ righe dove ogni riga è una combinazione univoca di SKU-PDV e il numero complessivo degli SKU è circa 532.

Invece il file di input DWH contiene $2 \cdot 10^6$ righe.

Per ogni riga sono disponibili le seguenti variabili: SKU, punto vendita (PDV), settimana di riferimento (WEEK), nome evento promozionale, presenza evento promozionale, prezzo, sconto, venduto, tipo evento promozionale.

Risulta evidente che le informazioni contenute nella generica riga i -esima del file di input DWH sono univoche per combinazioni di SKU-PDV-WEEK.

Per ogni SKU si crea quindi un dataset temporaneo, isolando le righe del file DWH dove compare lo SKU da analizzare. Successivamente vengono esclusi dal dataset i PDV per cui il singolo SKU non andrà nella promozione volantino presa in esame. Questa informazione è contenuta nel file di input Gestionale. Infatti, per esempio, può succedere che uno SKU nel passato sia stato venduto su 100 PDV ma la promozione attuale riguardi solo una parte di quei PDV, si ipotizza 60. Negli altri 40 PDV può accadere che lo SKU possa rientrare in un altro tipo di promozione (tra quelle elencate precedentemente), oppure possa essere venduto a prezzo regolare.

La grandezza di ogni singolo dataset, che dipende dalle caratteristiche della serie storica del singolo SKU, contiene il venduto per ogni combinazione di PDV in promozione - WEEK (compresa dal 2011 al 2017).

Si procede quindi ad una aggregazione dei dati di venduto per ottenere una serie storica aggregata su tutti i PDV presenti nella promozione per ogni singolo SKU.

In questo modo ad ogni SKU viene associata un'unica serie storica contenente tutti i PDV per cui quello SKU sarà in promozione.

Questa fase di aggregazione è molto importante come descritto nel capitolo 1 di questo elaborato per ottenere una previsione statisticamente più affidabile.

Va sottolineato il fatto che non tutti gli SKU presentano una serie storica affidabile¹⁵ e abbastanza lunga da poter essere utilizzata dai modelli di sales forecasting descritti nel capitolo 4. In particolare questi SKU possono essere modellati da modelli per il new product forecasting descritti in dettaglio nel paragrafo 4.7 di questo elaborato, a cui si rimanda.

Se consideriamo il caso specifico dell'offerta di volantino presa in esame va evidenziato che circa il 24% sul totale delle serie storiche non esiste oppure risulta troppo corta per essere utilizzata dai modelli di sales forecasting descritti nel capitolo 4 di questo elaborato.

Questo evidenzia come il mercato della GDO sia in continua evoluzione con l'introduzione di nuovi SKU per soddisfare le esigenze della clientela (per esempio con l'introduzione di nuovi gusti o fragranze) e i modelli per il new product forecasting ricoprano un ruolo non comprimario rispetto al problema della previsione della domanda.

Tenuto conto dello scarto appena descritto, si prendono in considerazione solo le serie storiche dei singoli SKU che risultano affidabili, nel caso specifico rimangono 404 SKU contro i 532 SKU iniziali.

5.3 Sales cleaning e sales forecasting applicati ai casi di studio aziendali

Le serie storiche individuate sono pronte per essere analizzate dal processo descritto in questo elaborato. In particolare saranno sottoposte a tutte le fasi del processo di sales cleaning (descritto nel capitolo 2).

Come prima fase le componenti di trend e stagionalità vengono calcolate utilizzando il metodo della scomposizione additiva (paragrafo 2.2).

A tal proposito si considera quindi un modello di regressione lineare con due variabili.

La prima indica la settimana di riferimento tenendo conto del totale della serie storica. Infatti ad ogni osservazione è associata una variabile indicante la settimana di riferimento **TIME** che assume quindi un valore progressivo dal minimo di 1 fino al massimo della lunghezza totale delle osservazioni. Per quanto riguarda la seconda, essa viene associata alla stagionalità **SEAS** ed è una variabile categoriale con 52/53 modalità legata al numero della settimana nell'anno x. (A volte è 53 perché l'ultima settimana dell'anno contiene più di quattro giorni relativi al mese di dicembre).

Si utilizza quindi un modello di regressione lineare del tipo:

¹⁵ Una serie storica di domanda può essere affidabile se rispecchia alcuni criteri individuati dai Demand Planner. In questo caso una serie storica di domanda di uno SKU è affidabile se nell'ultimo anno mobile risulta almeno un venduto > 0 per evento promozionale.

$$Y = \alpha + \beta * TIME + \gamma * SEAS \quad (5.0)$$

dove Y sono le vendite solo della serie storica regolare.

Si utilizza poi una selezione delle variabili tramite stepwise utilizzando sia la direzione forward che backward. Va evidenziato che se anche solo una modalità della variabile SEAS è significativa allora la variabile SEAS non viene scartata dalla selezione delle variabili.

In particolare per calcolare il **trend** viene utilizzato il coefficiente β calcolato nel precedente modello e moltiplicato per la variabile TIME. Il trend così calcolato può essere utilizzato per detrendizzare la serie storica di domanda regolare, tramite una differenza.

Per quanto riguarda il calcolo della variabile **stagionalità** vengono utilizzati i coefficienti γ del modello di regressione visto in precedenza. Tuttavia questi coefficienti non sono relativi alla stagionalità di un particolare periodo ma rappresentano lo scostamento tra quel periodo e quello di riferimento (ovvero la modalità omessa nel calcolo del modello di regressione).

Dal momento che l'effetto della stagionalità deve ridursi entro l'anno di riferimento, per definizione di stagionalità, e i coefficienti calcolati, che risultano ancora grezzi, possono sommare a una quantità positiva se è soliti utilizzare un accorgimento, ovvero si sottrarre la media dei coefficienti di stagionalità ad ogni singolo coefficiente.

Si ottiene quindi una variabile di **stagionalità** che potrà essere eliminata dal venduto della serie storica originale (già detrendizzata) mediante differenza per ottenere la serie storica del regolare destagionalizzata cercata.

A questo punto per rimuovere gli outliers dalla serie di domanda, già destagionalizzata e detrendizzata, si utilizza il metodo dell'intervallo di confidenza nella sua versione dinamica che risulta più flessibile rispetto alla sua versione statica (paragrafo 2.3).

Per quanto riguarda il metodo dell'**impatto promozionale**, descritto solo per aspetti teorici nel paragrafo 2.5, si utilizza un modello non lineare (Nonlinear Least Squares), in generale costruito nel seguente modo:

$$Y = \alpha * \exp(\beta * X) \quad (5.1)$$

dove:

- Y , la variabile risposta, è un vettore con i valori della serie D_t^* (ovvero la serie detrendizzata e destagionalizzata)
- α è un vettore di parametri di scala
- β è un vettore di parametri legato alle esplicative della promozione
- X è una matrice di variabili esplicative che descrivono le varie promozioni a cui è soggetta la serie di domanda

Il metodo di stima dei parametri α e β è iterativo perché dati due valori iniziali di α e β , il modello cerca la convergenza ottimale utilizzando Gauss Newton, ovvero un algoritmo per risolvere problemi di minimi quadrati non lineari. Una volta ottimizzato, viene stimato l'impatto della promozione come:

$$\hat{\Delta} = \exp(\hat{\beta} * X) \quad (5.2)$$

Infatti se all'istante t non si è in presenza di promozioni allora $X=0$ e di conseguenza la stima dell'impatto promozionale corrisponderà ad 1.

Nel caso di studio specifico si vuole ottenere una previsione della domanda commerciale promozionale di un particolare tipo di promozione chiamata volantino. La 5.1, utilizzando le variabili esplicative a disposizione, viene modificata nel seguente modo:

$$Y = a * \exp(c * (-sconto\ unitario) + b * volantino) \quad (5.3)$$

dove:

- Y , la variabile risposta, è un vettore con i valori della serie D_t^* (ovvero detrendizzata e destagionalizzata)
- a è un parametro di scala
- b è il coefficiente legato al tipo di promozione
- c è il coefficiente di elasticità al prezzo
- sconto unitario è una variabile esplicativa definita come la differenza tra prezzo regolare e prezzo promozionale
- volantino è una variabile dicotomica che assume valore 1 quando è presente una promozione di tipo volantino mentre assume valore 0 negli altri casi.

Il metodo di stima dei parametri a, b, c è iterativo perché dati due valori iniziali di a, b, c il modello cerca la convergenza ottimale utilizzando Gauss Newton, ovvero un algoritmo per risolvere problemi di minimi quadrati non lineari.

Il modello fornisce quindi le stime dei coefficienti cercati \hat{b} e \hat{c} .

Una volta ottimizzato, viene stimato l'impatto della promozione $\hat{\Delta}_t$ per ogni periodo temporale passato t .

Va sottolineato che in ogni periodo t può essere presente una tra le seguenti tre situazioni:

- 1) nessuna promozione. In questo caso lo sconto unitario sarà 0 così come la variabile volantino. Di conseguenza $\hat{\Delta}_t = 1$, implicando che $D_t^* = D_t^C$. Questo è fondamentale in quanto stimare un impatto promozionale in periodi nei quali non è presente una promozione è concettualmente sbagliato e può portare a risultati predittivi distorti.

- 2) promozione diversa dal volantino. In questo caso lo sconto unitario è diverso da 0 mentre la variabile volantino = 0. Ciò implica che l'impatto della promozione sarà diverso da 0 e in formule risulta:

$$\widehat{\Delta}_t = \exp(\hat{c} * (-sconto\ unitario_t)) \quad (5.4)$$

dove \hat{c} è il coefficiente stimato precedentemente con il modello non lineare.

- 3) promozione di tipo volantino. In questo caso l'impatto promozionale viene calcolato come:

$$\widehat{\Delta}_t = \exp(\hat{c} * (-sconto\ unitario_t) + \hat{b} * volantino_t) \quad (5.5)$$

dove \hat{c} e \hat{b} sono i coefficienti stimati precedentemente con il modello non lineare.

Come descritto nel capitolo 2 di questo elaborato la serie storica corretta dagli outliers promozionali tramite il metodo dell'impatto promozionale risulta quindi essere:

$$D_t^C = \frac{D_t^*}{\widehat{\Delta}_t} \quad (5.6)$$

Per ottenere la serie storica D_t^B , ovvero la serie di domanda base finale rettificata da tutti i possibili outliers siano essi imputabili a un effetto di tipo promozionale o no, si aggiungono le componenti di trend e di stagionalità alla serie D_t^C .

Sarà quindi la serie D_t^B che verrà utilizzata come input dai modelli di *sales forecasting* descritti nel capitolo 4 di questo elaborato.

Questi modelli produrranno la previsione \widehat{y}_{t+k} , ovvero il forecast della domanda regolare per i periodi futuri t+k dove k=1,2,3,...

Tuttavia poiché l'obiettivo è quello di prevedere la domanda commerciale promozionale, che nel caso specifico sono i periodi k=1,2,3,4,5,6,7,8 viene stimato l'impatto promozionale $\widehat{\Delta}_{t+k}$ solo per quei periodi futuri. Per fare ciò si utilizzano le informazioni riguardanti le variabili sconto unitario e volantino contenute nel file di input Gestionale, mentre i coefficienti \hat{c} e \hat{b} sono quelli stimati nella 5.3. In formule risulta:

$$\widehat{\Delta}_{t+k} = \exp(\hat{c} * (-sconto\ unitario_{t+k}) + \hat{b} * volantino_{t+k}) \quad (5.7)$$

L'impatto promozionale stimato $\widehat{\Delta}_{t+k}$ viene quindi moltiplicato per la previsione baseline \widehat{y}_{t+k} dei rispettivi periodi t+k, dove k=1,2,3,4,5,6,7,8.

Tuttavia non tutti i 9 periodi riguardano la promozione volantino presa in esame. Infatti questa ha una durata di 15 giorni ciò significa che le prime 6 previsioni riguardano il periodo regolare dove gli impatti promozionali sono = 1, mentre le ultime 2 settimane riguardano il periodo promozionale cercato (dove gli impatti promozionali sono diversi da 0).¹⁶

Tradotto in formule si verifica la seguente situazione:

$$\widehat{y}_{t+k} = \begin{cases} \widehat{y}_{t+k} & \text{se } k = 1,2,3,4,5,6 \\ \widehat{y}_t * \widehat{\Delta}_t & \text{se } k = 7,8 \end{cases} \quad (5.8)$$

Va notato che l'obiettivo dei Demand Planner non è la stima delle previsioni nelle singole settimane, ma piuttosto la stima complessiva della domanda promozionale nell'intera promozione di volantino. Ciò è facilmente ottenibile dalla somma delle singole previsioni per le settimane 7 e 8.

Per valutare l'accuratezza dei vari modelli prodotti si utilizzeranno le 5 metriche di errore viste nel paragrafo 3.4. Tutte queste metriche si basano sui singoli errori e_t calcolati come $y_t - \widehat{y}_t$.

Questi errori vengono calcolati nella parte del dataset dedicato alla verifica, dove le quantità \widehat{y}_t vengono calcolate utilizzando il dataset di stima.

Come visto nel capitolo 3 di questo elaborato non si utilizzeranno solo questi due dataset (paragrafo 3.1) ma poiché l'obiettivo dei Demand Planner è la previsione della domanda commerciale promozionale e in particolare della promozione volantino si utilizzerà il metodo implementato ad hoc dello "one-step-promotional-ahead forecast" descritto nel paragrafo 3.3.

Se il numero di promozioni di volantino nella promozione sono n ciò implica che ci saranno n dataset di stima e n dataset di verifica, dove su questi ultimi verranno calcolati $2n$ errori e_t , dovuti al fatto che la previsione sarà per due settimane, ovvero rispettivamente la numero 7 e 8.

5.4 Modelli di sales forecasting nei casi di studio

Vengono utilizzati 7 dei modelli di sales forecasting descritti nel capitolo 4 di questo elaborato, a cui si rimanda per la loro spiegazione generale.

- 1) Modello di regressione lineare:

$$Y = \alpha + \beta_1 * TIME + \beta_2 * SEAS$$

¹⁶ Va sottolineato il fatto che per produrre le previsioni del periodo promozionale (k=7,8) bisogna giocoforza produrre anche previsioni del periodo regolare (k=1,2,3,4,5,6). Ciò è necessario per vincoli di processo, in quanto la previsione deve essere disponibile in anticipo agli attori coinvolti nel processo promozionale dell'azienda.

Le variabili TIME e SEAS sono descritte nel paragrafo 5.3, mentre la variabile risposta Y, è un vettore con i valori della serie D_t^B (ovvero la serie baseline). Come selezione di variabili si utilizza una stepwise minimizzando l'AIC.

- 2) Modello di regressione lineare con trasformata logaritmica della variabile risposta:

$$\log(Y) = \alpha + \beta_1 * TIME + \beta_2 * SEAS$$

Le variabili TIME e SEAS sono descritte nel paragrafo 5.3, mentre la variabile risposta, è il logaritmo di Y, ovvero un vettore con i valori della serie D_t^B (cioè la serie baseline).

Per selezionare la variabile si utilizza una stepwise minimizzando l'AIC.

- 3) ARIMA

Seguendo quanto riportato nel paragrafo 4.4.4 per ogni caso di studio si è studiato un particolare processo stocastico minimizzando anche in questo caso l'AIC.

Nel primo caso utilizzando l'autocorrelazione delle serie si è scelto un modello ARMA, imponendo che il trend e la stagionalità fossero nulli; nel secondo caso un modello ARIMA poiché era presente solo il trend; mentre nel terzo caso si è scelto un modello SARIMA, in quanto c'era la componente stagionale ma non quella di tendenza.

- 4) Croston:

Per tutti e tre i casi di studio viene utilizzato il modello di Croston descritto in dettaglio nel paragrafo 4.4.3 di questo elaborato.

- 5) Holt-Winters

Nei 3 casi di studio vengono utilizzati rispettivamente tre modelli di Holt-Winters diversi.

Nel primo caso poiché la serie è regolare e non c'è presenza di trend e stagionalità i coefficienti β e γ sono posti a zero. Nel secondo caso $\gamma = 0$ poiché non c'è stagionalità ma una componente di tendenza; mentre nell'ultimo caso succede il contrario, ovvero non c'è la stagionalità ma solo il trend e quindi viene posto $\beta = 0$.

Per la descrizione dettagliata si rimanda al paragrafo 4.4.2

- 6) Una rete neurale:

Una rete neurale per serie storiche, paragrafo 4.6, è stimata con valori ritardati di Y come input e un singolo strato latenti. Gli input sono i ritardi da 1 a p e i

ritardi da m a mP dove m è la frequenza della serie, che nei casi di studio è 53. Vengono stimate molte ripetizioni della rete, ciascuna con un peso casuale. Questi vengono poi mediati quando si calcolano le previsioni. La rete è addestrata per la previsione in un'unica fase. Le previsioni multi-step sono calcolate in modo ricorsivo. Per i dati non stagionali (ovvero i primi due casi di studio), il modello adattato è denotato come un modello NNAR (p, k), dove k è il numero di nodi nascosti. Quest'ultimo è analogo a un modello AR (p) ma con funzioni non lineari. Per l'ultimo dei tre casi di studio, dove c'è una presenza di stagionalità, il modello adattato è chiamato modello NNAR (p, P, k) [m], che è analogo a un modello ARIMA ($p, 0,0$) ($P, 0,0$) [m] ma con funzioni non lineari.

- 7) CF ovvero una combinazione pesata di tutti i precedenti 6 modelli. Il peso scelto è uniforme per ogni modello, secondo la definizione del paragrafo 4.8 di questo elaborato.

Per confrontare la previsione prodotta per l'offerta presa in esame di tipo volantino dai modelli di sales forecasting con il venduto reale si devono tenere presenti alcuni concetti.

Il primo riguarda il tipo di previsione elaborata. Infatti questa è a livello aggregato per SKU - PDV totali in promo - WEEK ma il forecast deve essere riproporzionato per i singoli PDV in modo da assegnare ad ognuno di essi delle quantità specifiche di ogni SKU. Questo è importante per la successiva fase di prenotazione dei singoli PDV. (vedasi figura 5.1)

Il secondo concetto importante è il fatto che il forecast segue un processo aziendale per cui la quantità relativa alla previsione non è mostrata direttamente al PDV ma ci sono n step successivi del processo che comportano approssimazioni e arrotondamenti legate a scelte manageriali/logistiche.

Questa è la causa principale per il quale il confronto diretto tra la previsione e il consuntivo (ovvero il venduto) non risulta abbastanza soddisfacente. Ciò si noterà con i prossimi tre casi di studio che verranno presentati.

5.5 Caso di studio 1

Il motivo di aver scelto questa prima serie storica è il fatto che è essenzialmente regolare e che presenta un buon numero di promozioni passate. Anche se bisogna tenere conto che le situazioni come quella appena descritta in realtà non sono molte.

La figura 5.2 mostra la serie storica del prezzo regolare (più in alto) e quella dello sconto (più in basso). Si nota come entrambe le serie, per definizione simmetriche rispetto all'asse delle ascisse, siano stazionarie.

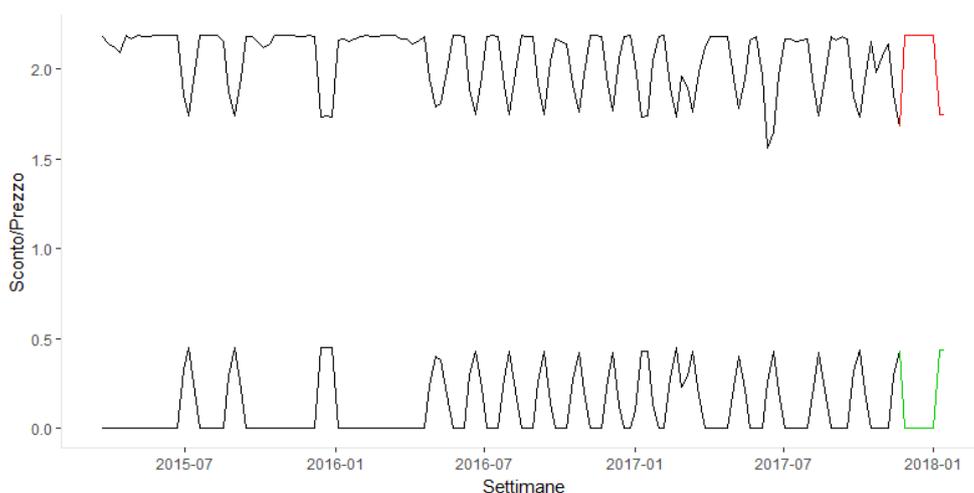


Figura 5.2

Le settimane in cui è richiesta la previsione della domanda promozionale del volantino sono la seconda e la terza settimana di gennaio. Esse iniziano rispettivamente i giorni 08/01/18 e 15/01/18. In altre parole in queste due settimane la promozione di volantino risulta essere attiva nei PDV.

Lo SKU ha 32 PDV nel file di input DWH in cui ci sono quindi 4171 righe.

Confrontando i dati del file di input Gestionale si nota come i PDV per cui lo SKU andrà in promozione siano solo 31 PDV, quindi il file di input DWH viene filtrato solo per questi PDV e risulta avere 4166 righe.

La figura 5.3 mostra la serie originale di partenza dello SKU preso in esame. Si sono evidenziate le promozioni di volantino con dei pallini pieni, mentre le altre promozioni con dei quadrati vuoti.

In totale nel passato ci sono quindi 12 promozioni di tipo volantino.

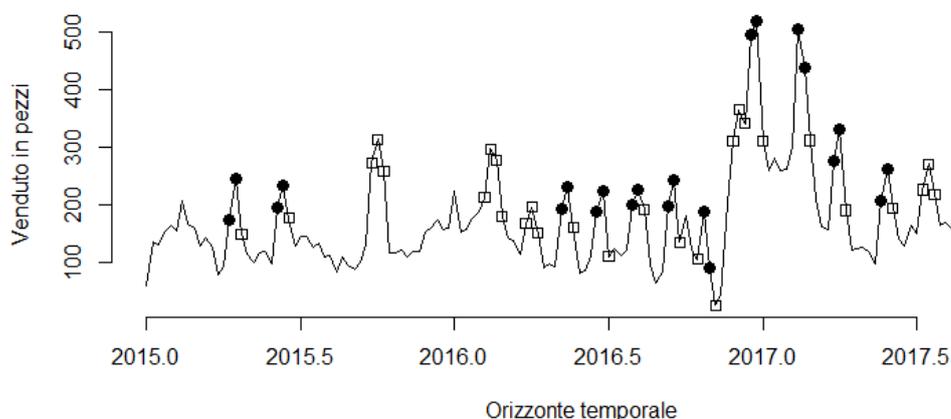


Figura 5.3

La figura 5.4 mostra la serie storica dove solamente la parte di vendite regolari è stata corretta seguendo i passaggi descritti nel capitolo 2. Utilizzando la stessa nomenclatura di tale capitolo la serie viene chiamata D_t^{PARZ} .

Si può notare come i picchi promozionali siano identici alla figura 5.3, mentre la settimane di vendite regolare cambiano, in particolare la settimana iniziale del 2017.

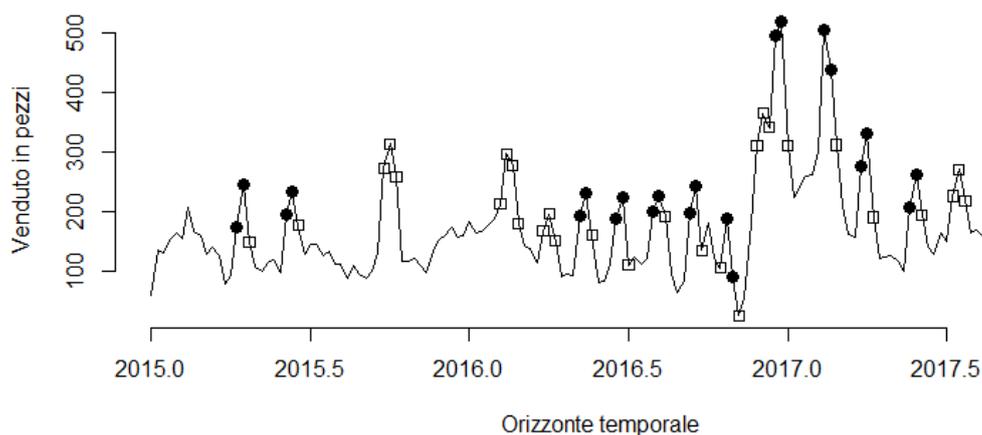


Figura 5.4

Alla serie della figura 5.4 viene quindi applicato il metodo dell'impatto promozionale descritto nel paragrafo 2.5 di questo elaborato.

Nella figura 5.5 vengono prese in considerazione solo le settimane dove è presente la promozione di volantino per evidenziare come i coefficienti di impatto promozionale seguano lo stesso andamento della serie della variabile sconto unitario che viene utilizzata per la loro costruzione nel metodo

dell'impatto promozionale. Infatti nel seguente grafico le due serie hanno un andamento pressoché identico, tranne per il fatto che l'asse y cambia. Il range della variabile sconto unitario è $(0,1)$ mentre quello dei coefficienti promozionali ha dominio potenzialmente $(0,+\infty)$.

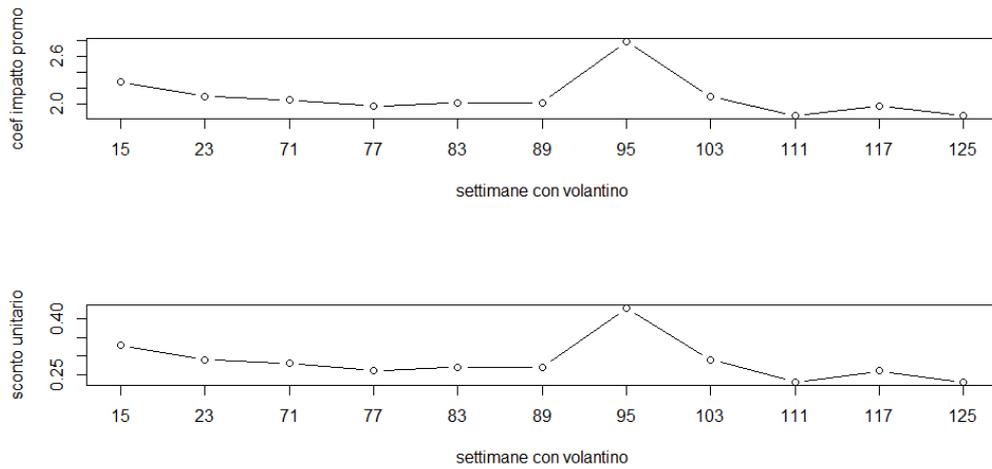


Figura 5.5

La serie finale (che nel capitolo 2 viene chiamata serie base di domanda D_t^B) risultante dal processo di sales cleaning che i modelli di sales forecasting utilizzano per la previsione della domanda commerciale è rappresentata nel seguente grafico.

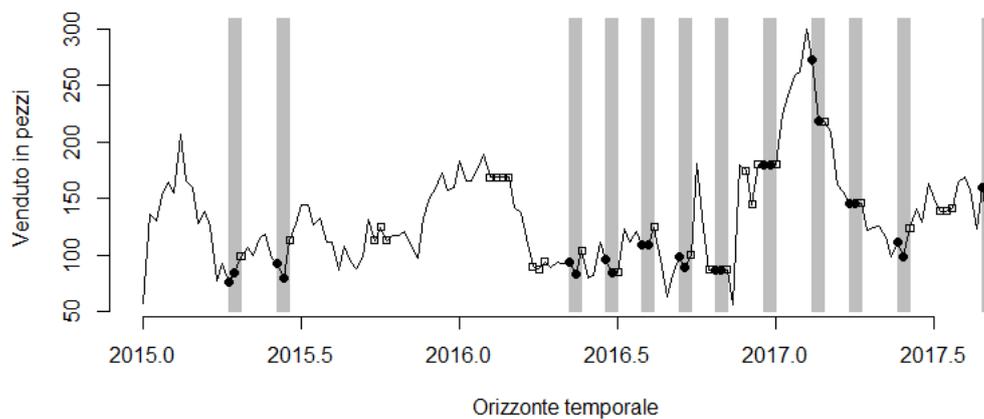


Figura 5.6

Le settimane in cui sono presenti promozioni di volantino sono evidenziate dalle bande grigie. Per completezza di informazione vengono riportate le posizioni delle prime settimane di ogni volantino: 15, 23, 71, 77, 83, 89, 95, 103, 111, 117, 125, 139.

La lunghezza delle osservazioni dei dataset di stima utilizzati nella stima modelli di sales forecasting con il metodo "one step promotional ahead forecasting" deve essere maggiore di 53 poiché i modelli utilizzano la

stagionalità (ovvero una variabile factor con 52/53 modalità) essendo i dati settimanali.

Per questa ragione vengono quindi utilizzate solo le promozioni di volantino che hanno la prima settimana coincidente rispettivamente all'osservazione in posizione 71, 77, 83, 89, 95, 103, 111, 117, 125 e 139. Di conseguenza i dataset di stima utilizzati sono 10 con numero di osservazioni rispettivamente di 70, 76, 82, 88, 94, 102, 110, 116, 124 e 138.

La tabella 5.1 indica gli impatti delle previsioni rispettivamente per la prima e per la seconda settimana (si ricorda che le previsioni richieste riguardano le settimane promozionali ovvero le settimana 7 e 8, in quanto le prime 6 sono di vendite regolari).

Intervallo	Impatto previsione (settimana 7)	Impatto previsione (settimana 8)
1	2,33	3,39
2	1,76	2,67
3	1,74	2,49
4	1,67	2,39
5	2,27	2,27
6	1,57	2,06
7	1,67	2,42
8	1,90	2,72
9	1,81	2,69
10	2,04	2,73

Tabella 5.1

A questo punto la serie baseline rappresentata nella figura 5.6 viene utilizzata in primo luogo dai modelli di sales forecasting per ottenere le previsioni cercate. Dall'altro dal metodo dello "one step promotional ahead" per calcolare gli errori negli insiemi di verifica.

In particolare la tabella 5.2 mostra gli errori e_t calcolati utilizzando il metodo dello "one step promotional ahead" (descritto nel paragrafo 3.3 di questo elaborato) per ognuno dei modelli di sales forecasting utilizzato.

Intervallo	Week	Reg. lineare	Reg. log lineare	ARIMA	Croston	Holt-Winters	Rete Neurale
1	1	-98,38	-5,92	-18,62	-34,82	-26,92	-35,64
	2	-193,17	-10,80	-77,33	-113,62	-92,78	-105,95
2	3	-27,79	-3,16	9,94	-8,60	-12,20	4,47
	4	-105,06	-7,34	-47,78	-79,03	-110,97	-46,65
3	5	-14,01	-3,01	-3,05	-10,92	-22,18	2,97
	6	-80,09	-6,48	-64,38	-74,97	-107,18	-38,52
4	7	-4,68	-2,68	68,13	50,11	85,23	34,22
	8	-47,41	-5,89	56,56	18,39	97,88	3,49
5	9	-88,00	-5,60	-31,36	-27,18	-6,59	-46,14
	10	-186,00	-6,33	-129,36	-139,02	-72,84	-146,38
6	11	303,41	-1,28	230,05	242,76	196,12	231,19
	12	264,38	-3,61	167,77	202,10	96,94	198,08
7	13	262,76	-1,80	31,51	36,50	-20,75	-0,16
	14	88,46	-5,54	-246,18	-218,37	-369,95	-295,62
8	15	-23,50	-3,80	-18,47	-20,51	22,40	-17,68
	16	-101,88	-7,74	-93,41	-95,94	18,93	-32,86
9	17	-70,31	-3,62	26,07	24,57	40,98	20,49
	18	-153,75	-7,75	-9,50	-10,69	36,45	-22,44
10	19	32,85	-4,31	83,98	83,18	96,69	87,99
	20	-91,59	-7,80	-22,35	-21,21	18,64	13,72

Tabella 5.2

Gli errori e_t della tabella 5.2 vengono utilizzati per il calcolo delle 5 metriche di errore principali per ciascun modello di sales forecasting mostrate nella tabella 5.3.

Modelli	ME	MPE	MAE	MAPE	RMSE
Regressione lineare	-16,69	-23,9	111,87	43,31	141,04
Regressione log lineare	-5,22	-95,18	5,22	95,18	5,72
ARIMA	-4,39	-8,50	71,79	26,99	99,62
Croston	-9,86	-12,30	75,62	28,70	103,83
Holt-Winters	-6,61	-5,90	77,63	28,20	112,85
Rete Neurale	-9,57	-11,65	69,23	26,35	107,53

Tabella 5.3

A questo punto tutti i modelli di sales forecasting vengono applicati sull'intera serie di figura 5.6, ovvero sulle 140 osservazioni, per ottenere le previsioni cercate.

Dalla tabella 5.3 i Demand Planner hanno il compito di decidere quale sarà il modello di sales forecasting da utilizzare per risolvere il problema previsionale.

In questo caso si farà riferimento esclusivamente al MAPE, anche se i Demand Planner possono scegliere altre metriche o la combinazione di queste (vedasi paragrafo 3.4). Nello specifico il MAPE minore è quello relativo alla Rete Neurale, tuttavia anche il valore del MAPE relativo all'ARIMA non si discosta molto.

Per questo motivo vengono analizzati i residui di entrambi i modelli di sales forecasting per valutare quale poter scegliere.

Il correlogramma della serie, mostrato nella figura 5.7, è di utilità per la scelta dei parametri dell'ARIMA. In questo caso il modello è un ARIMA (1,0,0) poiché nell'autocorrelogramma solo al primo ritardo c'è un valore superiore alle bande di confidenza al 95%, mentre nel correlogramma parziale nessun ritardo cade al di fuori delle bande di confidenza al 95%.

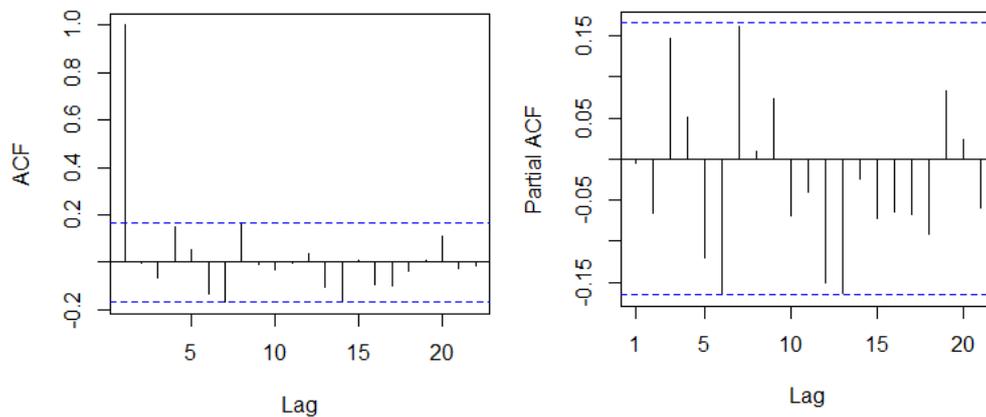


Figura 5.7

Analizzando i residui del modello ci si accorge che essi seguono abbastanza un processo White Noise. Infatti solo tre residui oltrepassano le bande di confidenza al 95%. Di conseguenza il modello ARIMA (1,0,0) è un buon candidato per la serie in esame.

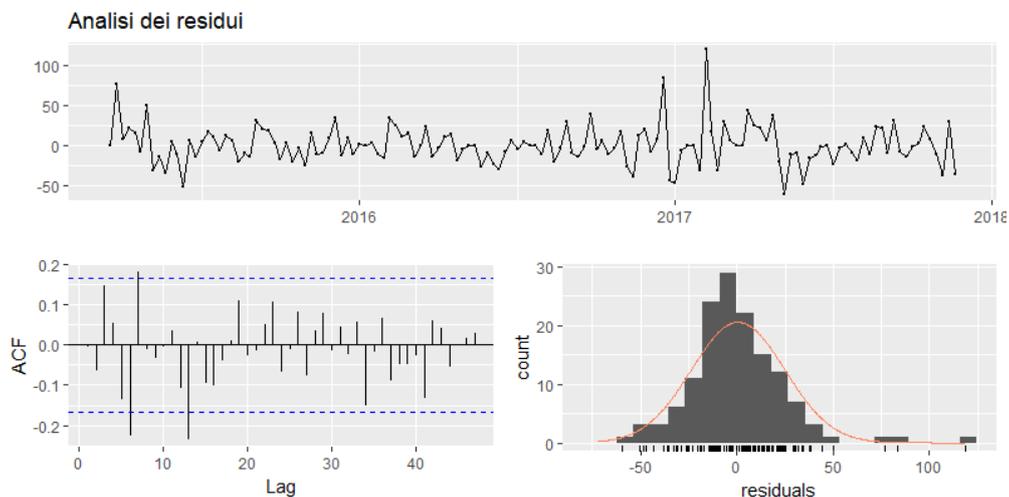


Figura 5.8

Per quanto riguarda l'analisi dei residui della rete neurale NNAR(7,1), dalla figura 5.9 si può notare come tutti i residui siano contenuti nelle bande di confidenza al 95%. Proprio a causa di questo la rete neurale viene scelta come miglior modello per il problema previsionale in esame.

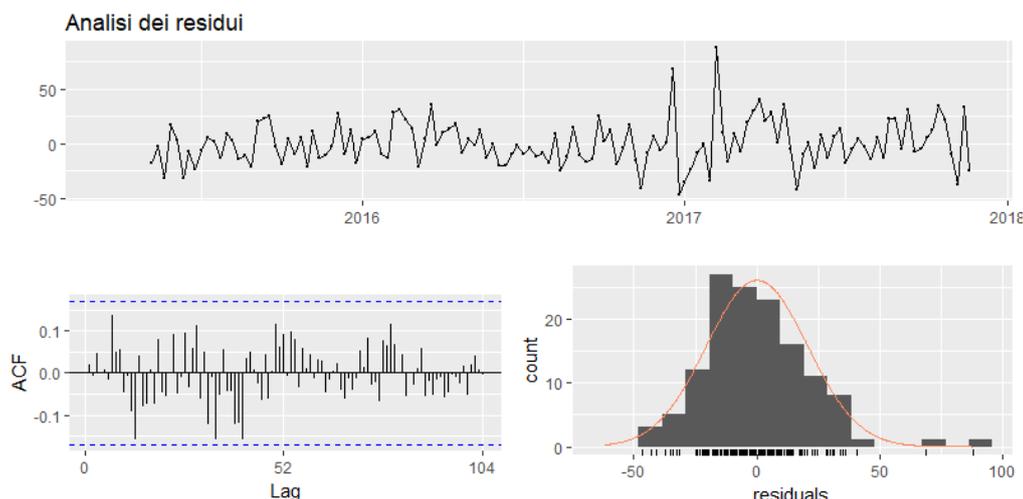


Figura 5.9

Vengono infine presentate le previsioni puntuali in pezzi di tutti i modelli presi in esame con l'evidenziazione del modello scelto che in questo caso è la rete neurale. Inoltre l'ultima colonna evidenzia la previsione in colli delle settimane promozionali (7 e 8). Si nota come la previsione in colli per la rete neurale non differisce molto dalla previsione del modello ARIMA visto in precedenza. Mentre la previsione in colli della combinazione di tutti i modelli (CF) è esattamente uguale con quella della rete neurale scelta.

Modelli	Settimane di previsione								Colli Promo
	1	2	3	4	5	6	7	8	
Regressione lineare	153	153	154	154	154	154	439	440	74
Regressione log lineare	145	145	145	146	146	146	416	417	70
ARIMA	108	113	111	109	107	115	322	316	54
Croston	122	122	122	122	122	122	347	347	58
Holt-Winters	116	108	99	91	82	74	184	159	29
Rete Neurale	125	124	124	124	124	124	352	352	59
CF	NA	NA	NA	NA	NA	NA	NA	NA	59

Tabella 5.4

Per completezza di informazione si riporta la figura 5.10 in cui si vede la serie storica originale sovrapposta alla serie stimata della rete neurale, ovvero il modello scelto, rappresentata da una linea tratteggiata. Le promozioni sono

evidenziate da pallini pieni mentre la previsione si trova alla fine della serie con gli ultimi 2 valori, ovvero le settimane 7 e 8.

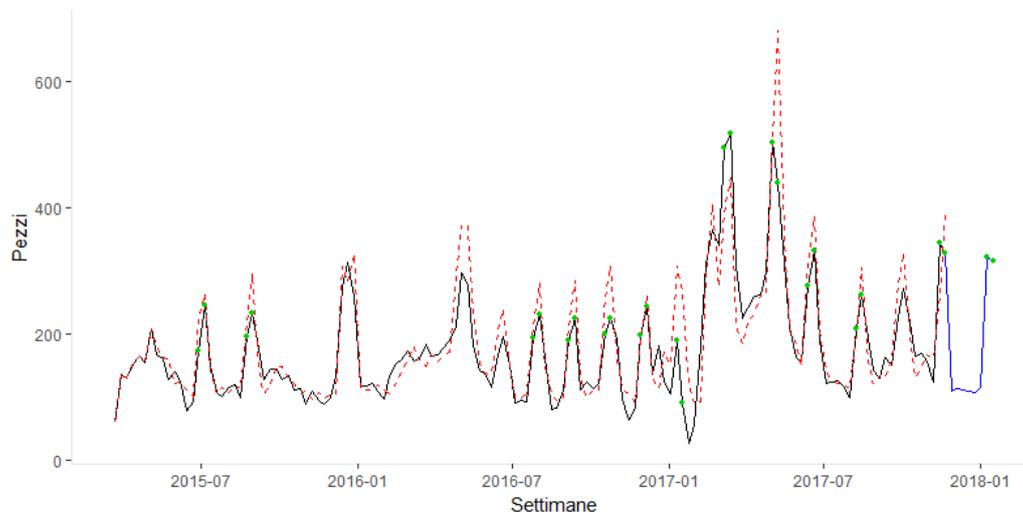


Figura 5.10

La previsione di 59 colli è stata successivamente spaccata per ognuno dei 31 PDV per cui il prodotto andrà in promozione. Per fare ciò si è scelto di riproporzionarla secondo il fatturato dei rispettivi negozi.

Così facendo sono stati allocati rispettivamente 5 colli a 2 PDV, 4 colli a 4 PDV, 3 colli a 6 PDV, 2 colli a 10 PDV e 1 collo a 8 PDV. Dall'allocazione restano esclusi 5 PDV a cui non viene fornito neanche un collo.

Tuttavia le scelte manageriali impongono che ogni PDV debba ricevere almeno un collo se il prodotto è in promozione. Questo è il caso degli ultimi 5 PDV citati. Per questa ragione la previsione finale passa da 59 colli a 64 colli.

Analizzando il consuntivo aggregato dei PDV, questo prodotto a fronte di una previsione di 59 colli effettiva, arrotondata a 64, ha venduto 55 colli.

5.5 Caso di studio 2

Il motivo di aver scelto questa serie storica è il fatto che presenta un trend decrescente e che possiede un buon numero di promozioni passate, dovute anche al fatto che il prodotto non è nuovo (infatti viene venduto dal 2011).

La figura 5.11 mostra la serie storica del prezzo regolare (più in alto) e quella dello sconto (più in basso). Si nota come entrambe le serie, per definizione simmetriche rispetto all'asse delle ascisse, siano stazionarie, anche se nel 2011 il prezzo regolare ha subito delle variazioni. Ciò significa molto probabilmente che il prezzo di uscita sul mercato era troppo alto ed è quindi stato rivisto al ribasso.

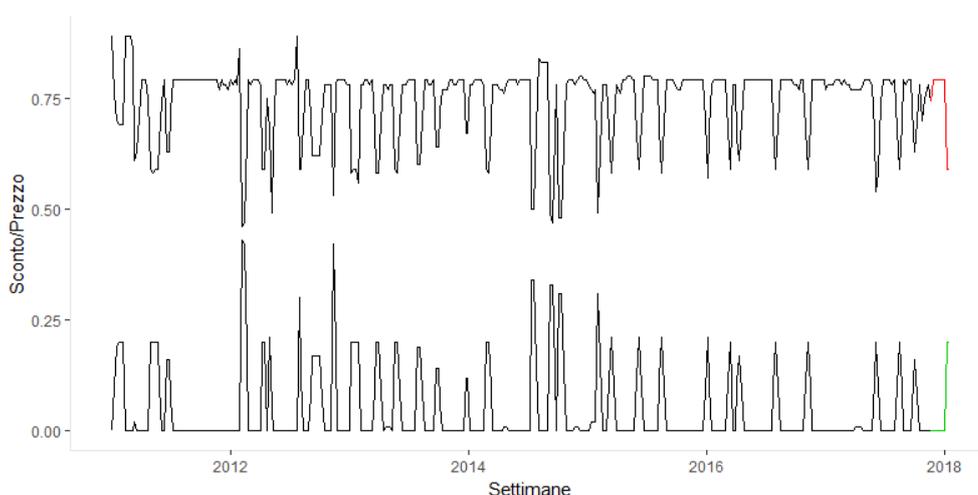


Figura 5.11

Le settimane in cui è richiesta la previsione della domanda promozionale del volantino sono la seconda e la terza settimana di gennaio. Esse iniziano rispettivamente i giorni 08/01/18 e 15/01/18. In altre parole in queste due settimane la promozione di volantino risulta essere attiva nei PDV.

Lo SKU ha 32 PDV nel file di input DWH in cui ci sono quindi 10.644 righe.

Confrontando i dati del file di input Gestionale si nota come i PDV per cui lo SKU andrà in promozione siano solo 30 PDV, quindi il file di input DWH viene filtrato solo per questi PDV e risulta avere 10.640 righe.

La figura 5.12 mostra la serie originale di partenze dello SKU preso in esame. Si sono evidenziate le promozioni di volantino con dei pallini pieni, mentre le altre promozioni con dei quadrati vuoti. In totale nel passato ci sono quindi 21 promozioni di tipo volantino.

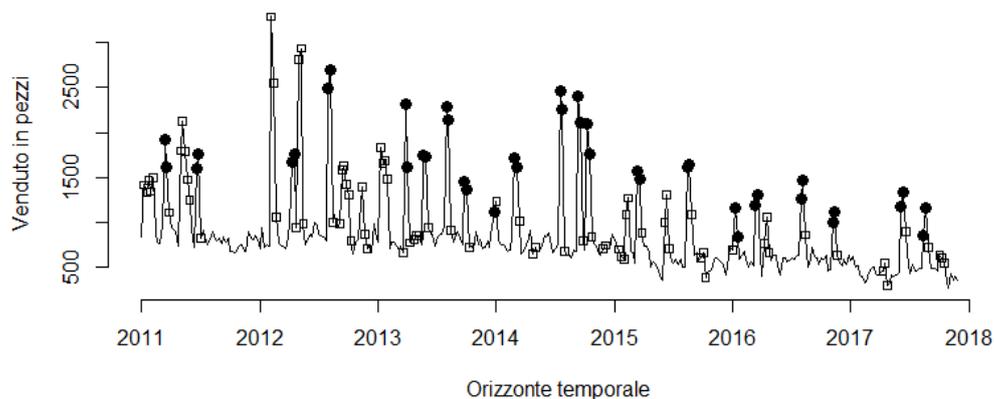


Figura 5.12

La figura 5.13 mostra la serie storica dove solamente la parte di vendite regolari è stata corretta seguendo i passaggi descritti nel capitolo 2. Utilizzando la stessa nomenclatura di tale capitolo la serie viene chiamata D_t^{PARZ} .

Si può notare come i picchi promozionali siano identici alla figura 5.12, mentre la settimane di vendite regolare cambiano, anche se dalla figura non si riesce a notarlo a causa della notevole differenza tra le vendite in regolare e in promozionale.

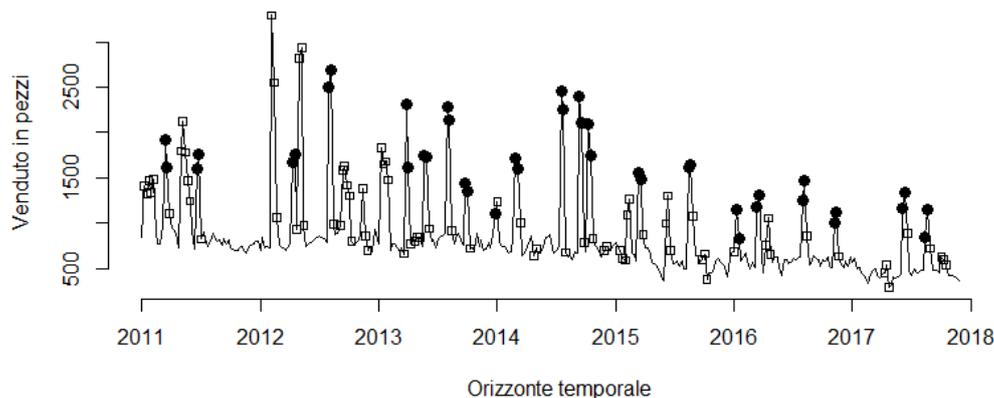


Figura 5.13

Alla serie della figura 5.13 viene quindi applicato il metodo dell'impatto promozionale descritto nel paragrafo 2.5 di questo elaborato.

Nella figura 5.14 vengono prese in considerazione solo le settimane dove è presente la promozione di volantino per evidenziare come i coefficienti di impatto promozionale seguano lo stesso andamento della serie della variabile

sconto unitario che viene utilizzata per la loro costruzione nel metodo dell'impatto promozionale. Infatti nella seguente figura le due serie hanno un andamento pressoché identico, tranne per il fatto che l'asse y cambia. Il range della variabile sconto unitario è (0,1) mentre quello dei coefficienti promozionali ha dominio potenzialmente (0, +∞).

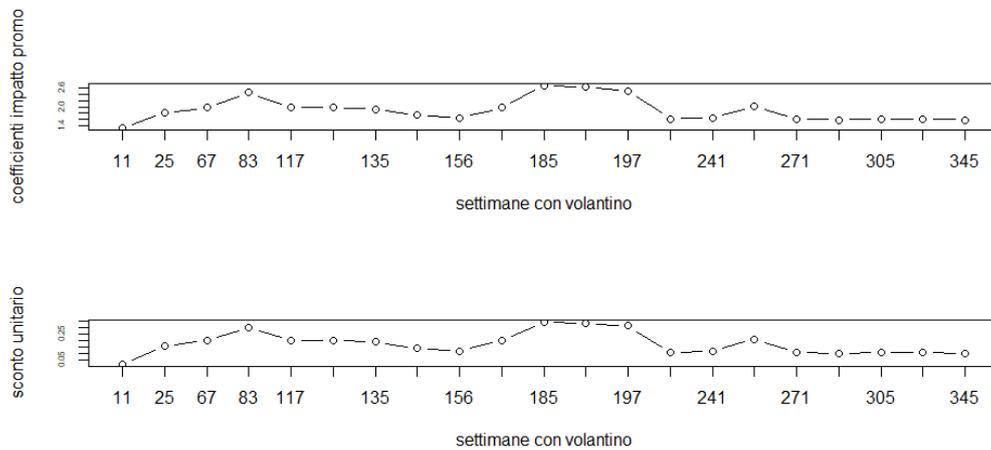


Figura 5.14

La serie finale (che nel capitolo 2 viene chiamata serie base di domanda D_t^B) risultante dal processo di sales cleaning che i modelli di sales forecasting utilizzano per la previsione della domanda commerciale è rappresentata nella seguente figura.

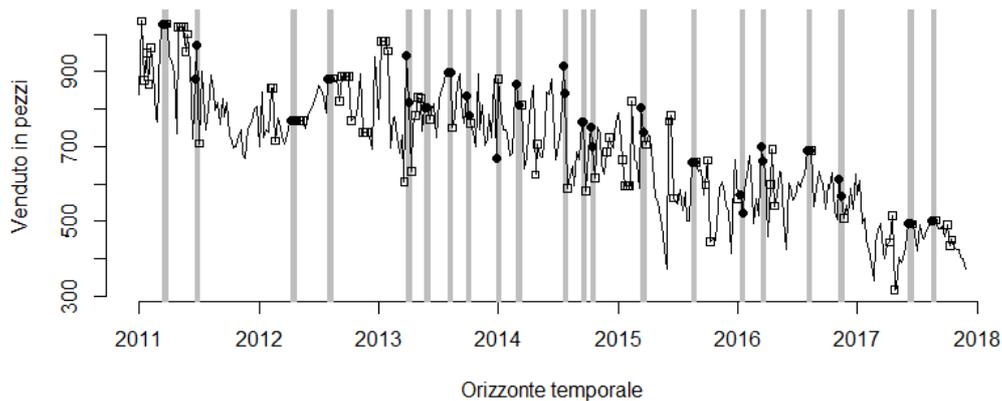


Figura 5.15

Le settimane in cui sono presenti promozioni di volantino sono evidenziate dalle bande grigie. Per completezza di informazione vengono riportate le posizioni delle prime settimane di ogni volantino: 11, 25, 67, 83, 117, 125, 135, 143, 156, 165, 185, 193, 197, 219, 241, 262, 271, 291, 305, 335, 345.

La lunghezza delle osservazioni dei dataset di stima utilizzati nella stima modelli di sales forecasting con il metodo "one step promotional ahead forecasting" deve essere maggiore di 53 poiché i modelli utilizzano la

stagionalità (ovvero una variabile factor con 52/53 modalità) essendo i dati settimanali.

Per questa ragione non vengono prese in considerazione le promozioni di volantino che hanno la prima settimana coincidente rispettivamente all'osservazione in posizione 11 e 25. Di conseguenza i dataset di stima utilizzati sono 19 con numero di osservazioni rispettivamente di 66, 82, 116, 124, 134, 142, 155, 164, 184, 192, 196, 218, 240, 261, 270, 290, 304, 334, 344.

La tabella 5.5 indica gli impatti delle previsioni rispettivamente per la prima e per la seconda settimana (si ricorda che le previsioni richieste riguardano le settimane promozionali ovvero le settimana 7 e 8, in quanto le prime 6 sono di vendite regolari).

Intervallo	Impatto previsione (settimana 1)	Impatto previsione (settimana 2)
1	3,48	3,48
2	2,97	1,78
3	2,36	2,36
4	2,34	2,34
5	2,28	2,28
6	2,05	2,05
7	1,92	1,38
8	2,32	2,32
9	3,31	3,31
10	3,07	3,07
11	2,84	2,84
12	1,71	2,17
13	1,73	2,13
14	2,12	1,65
15	1,67	2,04
16	1,62	2,03
17	1,66	2,03
18	1,64	2,00
19	1,60	1,99

Tabella 5.5

A questo punto la serie baseline rappresentata nella figura 5.15 viene utilizzata in primo luogo dai modelli di sales forecasting per ottenere le previsioni cercate. Dall'altro dal metodo dello "one step promotional ahead" per calcolare gli errori negli insiemi di verifica.

In particolare la tabella 5.6 mostra gli errori e_t calcolati utilizzando il metodo dello "one step promotional ahead" (descritto nel paragrafo 3.3 di questo elaborato) per ognuno dei modelli di sales forecasting utilizzato.

Intervallo	Week	Reg. lineare	Reg log lineare	ARIMA	Croston	Holt-Winters	Rete Neurale
1	1	-984,99	-15,44	-937,11	-856,64	-792,80	-982,28
	2	-918,25	-15,38	-812,84	-769,64	-696,95	-895,28
2	3	125,34	-11,78	63,77	126,58	313,00	85,41
	4	1287,07	-3,83	1237,78	1275,37	1390,12	1256,03
3	5	899,10	-7,92	718,03	805,93	493,02	568,37
	6	281,01	-8,28	27,03	114,93	-196,22	-181,02
4	7	-65,12	-8,10	-64,67	-19,10	-47,86	-71,69
	8	-81,65	-8,10	-79,67	-34,10	-61,14	-99,59
5	9	209,32	-7,42	340,97	338,51	524,15	360,63
	10	24,06	-7,49	198,97	196,51	383,51	289,72
6	11	-138,48	-6,33	-218,23	-142,56	-142,63	-149,28
	12	-202,89	-6,39	-305,23	-229,56	-228,68	-244,56
7	13	-517,45	-5,73	-408,26	-419,16	-367,98	-461,83
	14	42,42	-2,04	143,33	135,50	172,92	135,85
8	15	-106,37	-7,94	-30,58	-27,99	-46,86	-114,00
	16	-244,14	-8,01	-145,58	-142,99	-160,67	-210,45
9	17	66,42	-14,09	1,28	94,70	-10,50	-88,90
	18	-118,05	-14,17	-199,72	-106,30	-209,62	-303,12
10	19	390,29	-12,41	374,36	429,05	168,75	181,16
	20	97,39	-12,54	76,36	131,05	-127,16	-174,71
11	21	234,25	-11,06	199,83	263,35	40,74	56,90
	22	-102,04	-11,24	-145,17	-81,65	-302,25	-328,33
12	23	525,61	-3,85	455,42	506,36	371,34	393,41
	24	195,12	-6,91	45,86	145,94	-23,68	-67,70
13	25	756,76	-3,83	665,61	744,35	490,61	640,47
	26	614,61	-6,38	457,60	576,79	266,94	424,72
14	27	-98,02	-6,51	-30,86	-47,46	-129,59	-61,13
	28	-157,46	-3,87	-96,74	-108,99	-171,29	-118,10
15	29	315,65	-3,57	272,68	318,87	193,31	241,44
	30	260,81	-5,87	166,91	250,36	98,85	110,40
16	31	244,63	-3,17	292,73	284,51	330,03	274,32
	32	189,18	-5,61	273,42	251,43	310,75	250,21
17	33	150,95	-3,61	62,47	145,10	64,44	63,22
	34	105,48	-5,85	-48,12	75,57	-20,98	-42,49
18	35	477,40	-3,14	470,07	494,77	341,74	473,93
	36	495,75	-5,24	495,88	519,26	335,12	533,51
19	37	71,89	-3,18	98,71	100,10	66,21	79,23
	38	181,62	-5,30	237,45	224,55	184,79	216,45

Tabella 5.6

Gli errori e_t della tabella 5.6 vengono utilizzati per il calcolo delle 5 metriche di errore principali per ciascun modello di sales forecasting mostrate nella tabella 5.7.

Modelli	ME	MPE	MAE	MAPE	RMSE
Regressione lineare	118,61	6,41	315,18	19,60	434,76
Regressione log lineare	-7,41	-99,61	7,41	99,61	8,26
ARIMA	101,41	5,50	286,82	17,69	398,31
Croston	146,40	8,24	303,57	18,75	412,34
Holt-Winters	73,78	3,40	270,45	16,37	374,28
Rete Neurale	53,71	2,79	295,52	18,18	401,50

Tabella 5.7

A questo punto tutti i modelli di sales forecasting vengono applicati sull'intera serie di figura 5.15 per ottenere le previsioni cercate.

Dalla tabella 5.7 i Demand Planner hanno il compito di decidere quale sarà il modello di sales forecasting da utilizzare per risolvere il problema previsionale. In questo caso si farà riferimento esclusivamente al MAPE, anche se i Demand Planner possono scegliere altre metriche o la combinazione di queste (vedasi paragrafo 3.4). Nello specifico il MAPE minore è quello relativo al modello di Holt-Winters, dove $\gamma = 0$ in quanto non c'è stagionalità.

Vanno tuttavia analizzati i residui di tale modello di sales forecasting per valutare se è un buon modello per la serie analizzata. In questo caso si nota che i residui sono contenuti dentro la banda di confidenza del 95% e quindi il modello di Holt-Winters può essere scelto per risolvere il problema previsionale in esame.

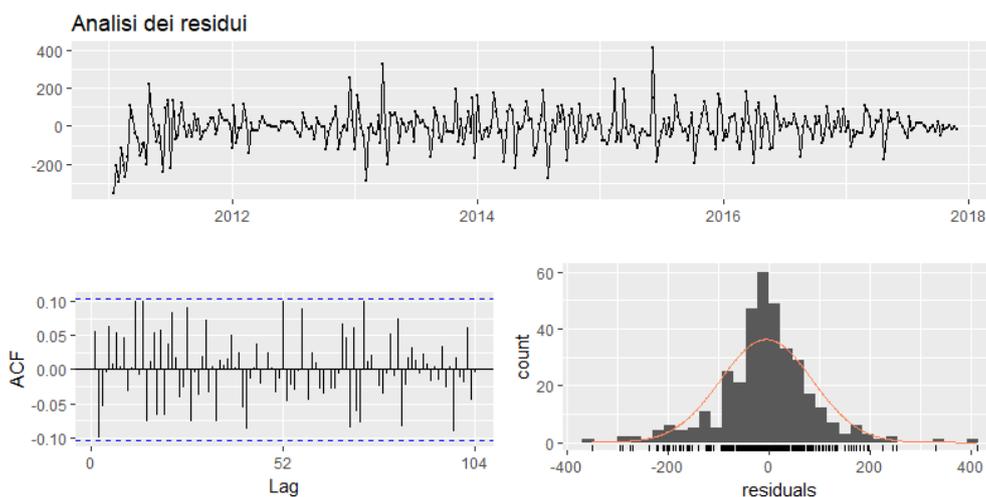


Figura 5.16

Vengono infine presentate le previsioni puntuali in pezzi di tutti i modelli presi in esame con l'evidenziazione di quello scelto che in questo caso è il modello di Holt-Winters (302 colli previsti). Inoltre l'ultima colonna evidenzia la previsione in colli delle settimane promozionali (7 e 8).

Modelli	Settimane di previsione								Colli Promo
	1	2	3	4	5	6	7	8	
Regressione lineare	371	362	353	344	335	327	627	609	206
Regressione log lineare	475	474	474	473	472	471	928	926	309
ARIMA	413	427	428	430	430	430	849	849	283
Croston	375	375	375	375	375	375	740	740	247
Holt-Winters	466	465	463	462	461	460	905	903	302
Rete Neurale	409	421	422	422	423	426	852	860	286
CF	NA	NA	NA	NA	NA	NA	NA	NA	278

Tabella 5.8

Per completezza di informazione si riporta la figura 5.17 in cui si vede la serie storica originale sovrapposta alla serie stimata della rete neurale, ovvero il modello scelto, rappresentata da una linea tratteggiata. Le promozioni sono evidenziate da dei pallini pieni mentre la previsione si trova alla fine della serie con gli ultimi due valori, ovvero le settimane 7 e 8.

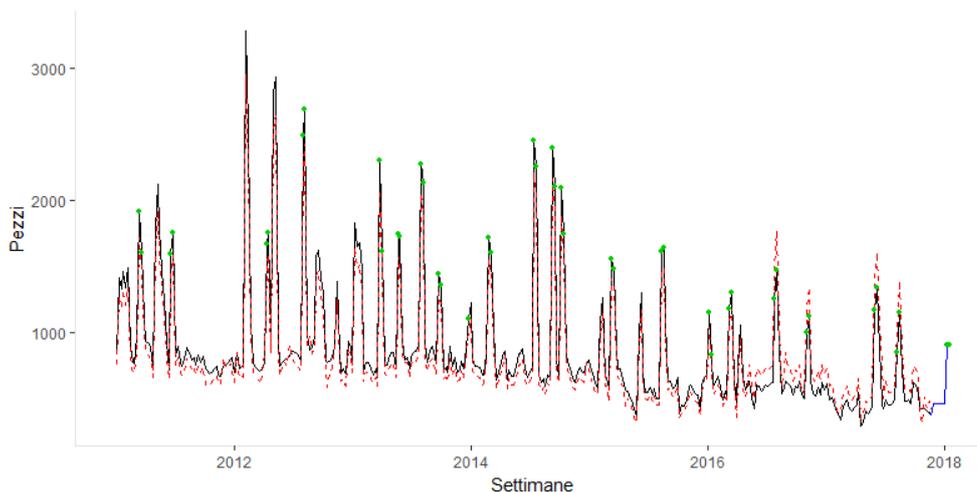


Figura 5.17

Anche in questo caso la previsione viene allocata nei 30 PDV per cui il prodotto andrà in promozione, ma questa volta (a differenza del caso di studio 1) a tutti i PDV viene assegnato almeno un collo.

Se si confronta il consuntivo con la previsione effettiva notiamo che c'è una differenza positiva di 50 colli, ovvero è stato venduto più di quello che è stato previsto. Ciò è spiegabile dal fatto che alcuni PDV hanno scelto una particolare collocazione della merce all'interno del negozio (in particolare è stato posto all'ingresso del negozio), in modo da offrire maggiore visibilità al prodotto. Questo ha portato ad aumentare le vendite e quindi durante la promozione i PDV hanno chiesto un'integrazione di colli dal magazzino centrale per evitare la rottura di stock del prodotto.

In pratica ciò significa che hanno rubato della scorte destinate ai PDV per le successive settimane (vendite regolari).

5.6 Caso di studio 3

Il motivo di aver scelto questa serie storica è il fatto che sembra presentare un trend decrescente oltre alla stagionalità. Inoltre possiede un buon numero di promozioni passate, dovute anche al fatto che il prodotto non è nuovo (viene venduto dal 2011).

La figura 5.18 mostra la serie storica del prezzo regolare (più in alto) e quella dello sconto (più in basso). Si nota come la serie del prezzo regolare sia molto altalenante, specie nell'ultimo periodo dove il prezzo regolare è molto più alto del solito. Ciò è dovuto al fatto che si volevano massimizzare i profitti per quel prodotto nelle ultime settimane della serie, ovvero quelle dove le vendite sono più alte dovute all'effetto della stagionalità. Questo trucco viene molto usato nella GDO e consiste in un aumento del prezzo regolare del bene per il periodo di massima vendibilità a cui viene applicato uno sconto molto importante che attira il cliente. Quindi, ad esempio, uno sconto che sembra essere del 50% in realtà potrebbe essere solo del 20.

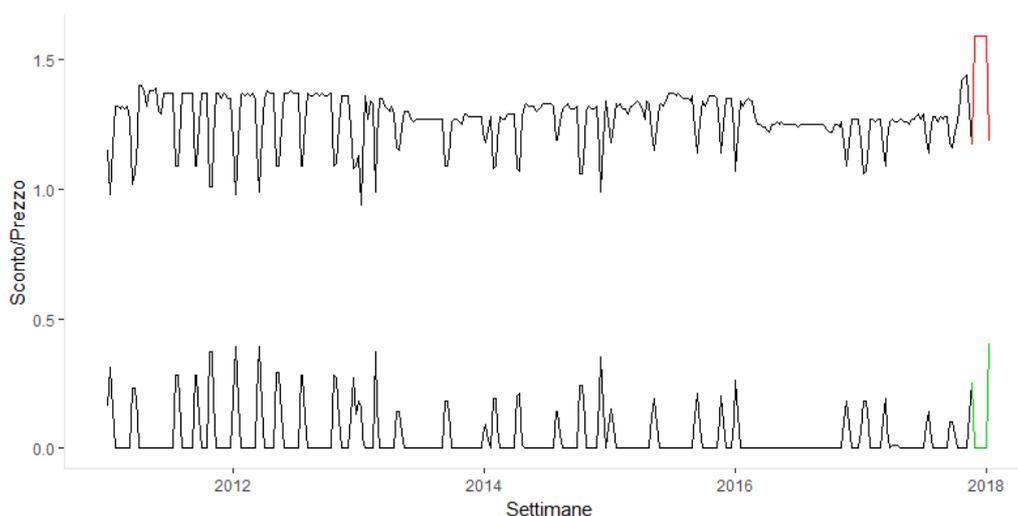


Figura 5.18

Le settimane in cui è richiesta la previsione della domanda promozionale del volantino sono la seconda e la terza settimana di gennaio. Esse iniziano rispettivamente i giorni 08/01/18 e 15/01/18. In altre parole in queste due settimane la promozione di volantino risulta essere attiva nei PDV.

Lo SKU ha 32 PDV nel file di input DWH in cui ci sono quindi 10.332 righe.

Confrontando i dati del file di input Gestionale si nota come i PDV per cui lo SKU andrà in promozione siano solo 30 PDV, quindi il file di input DWH viene filtrato solo per questi PDV e risulta avere 10.328 righe.

La figura 5.19 mostra la serie originale di partenze dello SKU preso in esame. Si sono evidenziate le promozioni di volantino con dei pallini pieni, mentre le altre promozioni con dei quadrati vuoti. In totale nel passato ci sono quindi 15 promozioni di tipo volantino.

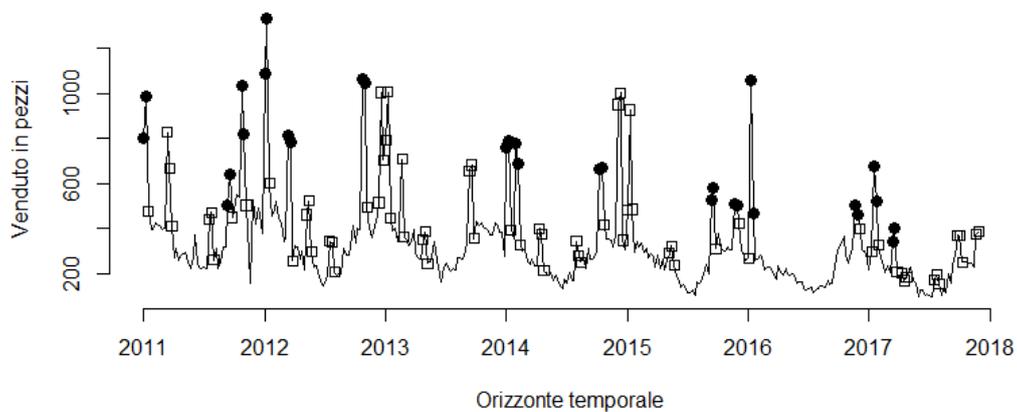


Figura 5.19

La figura 5.20 mostra la serie storica dove solamente la parte di vendite regolari è stata corretta seguendo i passaggi descritti nel capitolo 2. Utilizzando la stessa nomenclatura di tale capitolo la serie viene chiamata D_t^{PARZ} .

Si può notare come i picchi promozionali siano identici alla figura 5.19, mentre le settimane di vendite regolare cambiano, anche se dalla figura non si riesce a notarlo a causa della notevole differenza tra le vendite in regolare e in promozionale.

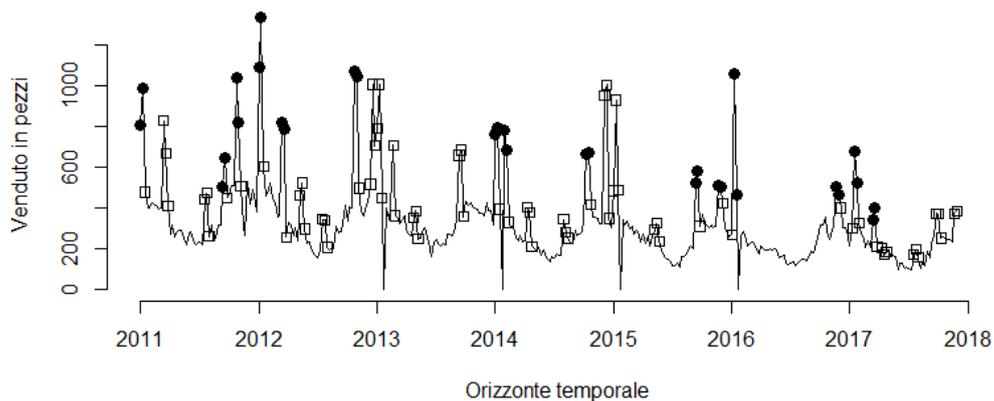


Figura 5.20

Alla serie della figura 5.20 viene quindi applicato il metodo dell'impatto promozionale descritto nel paragrafo 2.5 di questo elaborato.

Nella figura 5.21 vengono prese in considerazione solo le settimane dove è presente la promozione di volantino per evidenziare come i coefficienti di

impatto promozionale seguano lo stesso andamento della serie della variabile sconto unitario che viene utilizzata per la loro costruzione nel metodo dell'impatto promozionale. Infatti nella seguente figura le due serie hanno un andamento pressoché identico, tranne per il fatto che l'asse y cambia. Il range della variabile sconto unitario è (0,1) mentre quello dei coefficienti promozionali ha dominio potenzialmente (0, +∞).

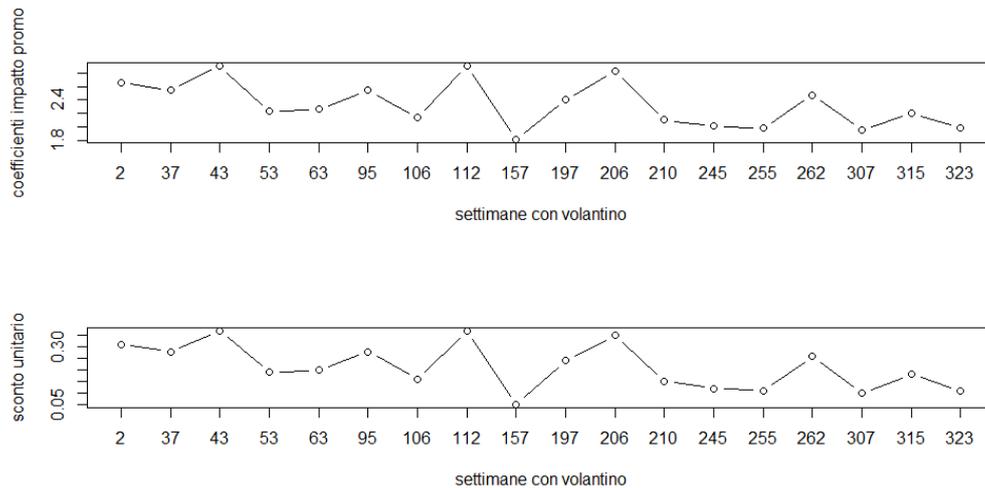


Figura 5.21

La serie finale (che nel capitolo 2 viene chiamata serie base di domanda D_t^B) risultante dal processo di sales cleaning che i modelli di sales forecasting utilizzano per la previsione della domanda commerciale è rappresentata nella seguente figura.

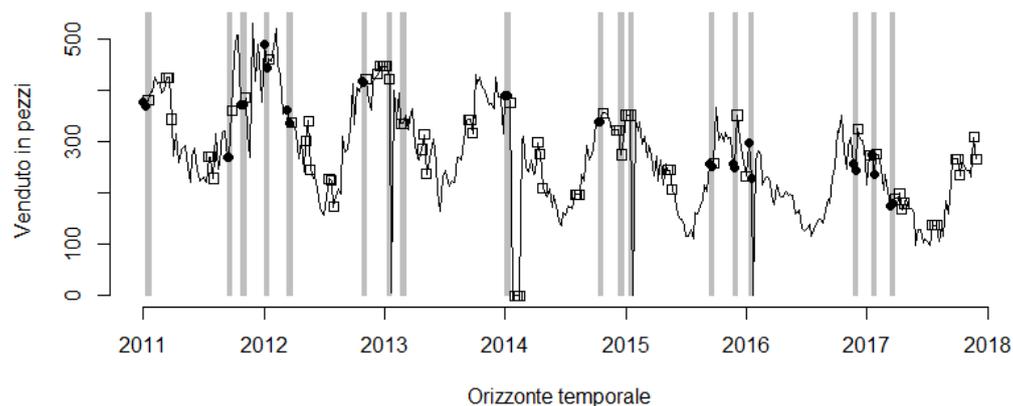


Figura 5.22

Le settimane in cui sono presenti promozioni di volantino sono evidenziate dalle bande grigie. Per completezza di informazione vengono riportate le posizioni delle prime settimane di ogni volantino: 2, 37, 43, 53, 63, 95, 106, 112, 157, 197, 206, 210, 245, 255, 262, 307, 315 e 323.

La lunghezza delle osservazioni dei dataset di stima utilizzati nei modelli di sales forecasting con il metodo “one step promotional ahead forecasting” deve essere maggiore di 53 poiché i modelli utilizzano la stagionalità (ovvero una variabile factor con 52/53 modalità) essendo i dati settimanali.

Per questa ragione non vengono prese in considerazione le promozioni di volantino che hanno la prima settimana coincidente rispettivamente all’osservazione in posizione 2, 37, 43, 53. Di conseguenza i dataset di stima utilizzati sono 14 con numero di osservazioni rispettivamente di 62, 94, 105, 111, 156, 196, 205, 209, 244, 254, 261, 306, 314, 322.

La tabella 5.9 indica gli impatti delle previsioni rispettivamente per la prima e per la seconda settimana (si ricorda che le previsioni richieste riguardano le settimane promozionali ovvero le settimana 7 e 8, in quanto le prime 6 sono di vendite regolari).

Intervallo	Impatto previsione (settimana 1)	Impatto previsione (settimana 2)
1	2,53	5,25
2	3,17	3,08
3	2,28	1,13
4	4,13	1,53
5	1,87	1,99
6	2,42	2,42
7	2,47	2,47
8	2,10	1,12
9	2,08	2,34
10	2,02	2,28
11	2,43	2,02
12	2,02	2,25
13	2,21	2,21
14	2,04	2,28

Tabella 5.9

A questo punto la serie baseline rappresentata nella figura 5.22 viene utilizzata in primo luogo dai modelli di sales forecasting per ottenere le previsioni cercate. Dall’altro dal metodo dello “one step promotional ahead” per calcolare gli errori negli insiemi di verifica.

In particolare la tabella 5.10 mostra gli errori e_t calcolati utilizzando il metodo dello “one step promotional ahead” (descritto nel paragrafo 3.3 di questo elaborato) per ognuno dei modelli di sales forecasting utilizzato.

Intervallo	Week	Reg. lineare	Reg. log lineare	ARIMA	Croston	Holt-Winters	Rete Neurale
1	1	-166,24	-7,98	-94,00	-97,02	NA	-96,95
	2	-1255,99	-23,73	-1098,68	-1104,9	NA	-1109,29
2	3	174,77	-10,70	-156,65	-151,60	NA	-183,78
	4	181,01	-10,20	-108,40	-138,23	NA	-157,47
3	5	266,28	-6,98	34,07	21,47	NA	37,25
	6	83,71	-0,51	-12,49	-37,34	NA	-17,90
4	7	-638,12	-17,03	-787,08	-786,48	-220,06	-873,38
	8	-138,33	-2,88	-187,87	-193,47	-28,70	-223,80
5	9	145,77	-4,12	117,18	134,35	-28,69	153,20
	10	135,68	-4,76	107,74	123,53	-85,26	151,38
6	11	-20,60	-7,27	337,49	474,00	-263,27	283,81
	12	-111,60	-7,39	210,03	383,00	-173,56	141,60
7	13	63,50	-7,08	-122,59	-98,81	36,98	-141,03
	14	70,11	-7,07	-117,59	-93,81	-80,16	-134,15
8	15	290,17	-8,65	92,80	115,04	131,99	102,48
	16	30,04	-1,13	-59,25	-49,26	74,64	-52,71
9	17	389,18	-4,82	224,90	244,10	62,37	216,94
	18	199,54	-0,03	111,31	121,56	45,05	109,43
10	19	61,09	-5,01	106,62	113,56	-135,23	76,93
	20	60,33	-6,30	110,07	117,87	-182,26	62,87
11	21	47,77	-4,74	-113,76	-97,88	318,34	-102,92
	22	-15,06	-6,15	-198,47	-180,55	-190,74	-178,90
12	23	500,09	-6,25	448,47	474,94	-219,02	467,74
	24	2,56	-4,85	-41,42	-19,38	-245,97	-29,30
13	25	11,48	-4,53	-97,79	-87,40	-24,74	-103,98
	26	-49,11	-5,84	-204,71	-193,13	-163,50	-187,86
14	27	86,65	-5,27	97,81	112,71	-316,56	82,45
	28	493,34	-5,53	-57,19	-42,29	-289,41	-61,29

Tabella 5.10

Dalla tabella 5.10 si nota come per i primi 2 intervalli non ci siano gli errori del modello Holt-Winters. Ciò è dovuto al fatto che per tenere conto della stagionalità il modello ha bisogno di almeno due anni di storico e quindi di almeno 106 osservazioni mentre i primi tre intervalli di stima considerati hanno rispettivamente 62, 94, 105 osservazioni.

Gli errori e_t della tabella 5.10 vengono utilizzati per il calcolo delle 5 metriche di errore principali per ciascun modello di sales forecasting mostrate nella tabella 5.11.

Modelli	ME	MPE	MAE	MAPE	RMSE
Regressione lineare	24,17	1,91	195,37	27,76	318,27
Regressione log lineare	-6,60	-101,04	6,60	99,04	7,97
ARIMA	-55,49	-11,23	188,72	28,50	292,10
Croston	-36,73	-8,29	199,14	29,53	306,14
Holt-Winters	-127,50	-26,91	185,71	34,13	264,91
Rete Neurale	-65,10	-12,59	190,83	28,47	300,89

Tabella 5.11

A questo punto tutti i modelli di sales forecasting vengono applicati sull'intera serie di figura 5.22 per ottenere le previsioni cercate.

Dalla tabella 5.11 i Demand Planner hanno il compito di decidere quale sarà il modello di sales forecasting da utilizzare per risolvere il problema previsionale. In questo caso si farà riferimento esclusivamente al MAPE, anche se i Demand Planner possono scegliere altre metriche o la combinazione di queste (vedasi paragrafo 3.4). Nello specifico il MAPE minore è quello relativo al modello di regressione lineare, ma anche quello della Rete Neurale non si discosta molto perciò si prenderà in considerazione anche questo modello.

Vanno analizzati i residui di tali modelli di sales forecasting per valutare quale è il migliore modello per la serie analizzata.

In questo caso si nota che i residui del modello di regressione lineare sono contenuti dentro la banda di confidenza del 95% e quindi il modello di regressione lineare può essere scelto per risolvere il problema previsionale in esame.

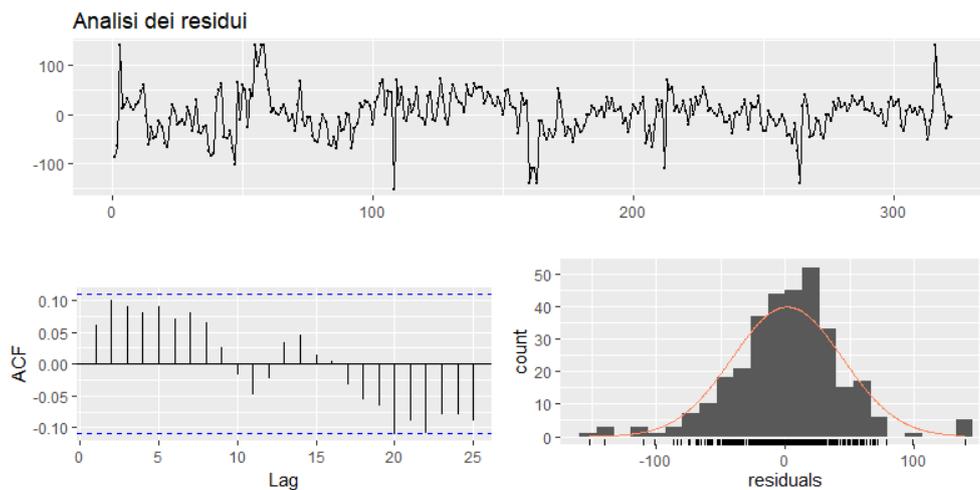


Figura 5.23

Nell'analisi dei residui di figura 5.24 si nota come alcuni residui superino le bande di confidenza al 95%. Ciò significa che il modello NNAR(5,1,1)[52] non è il miglior modello per risolvere il problema previsionale in esame.

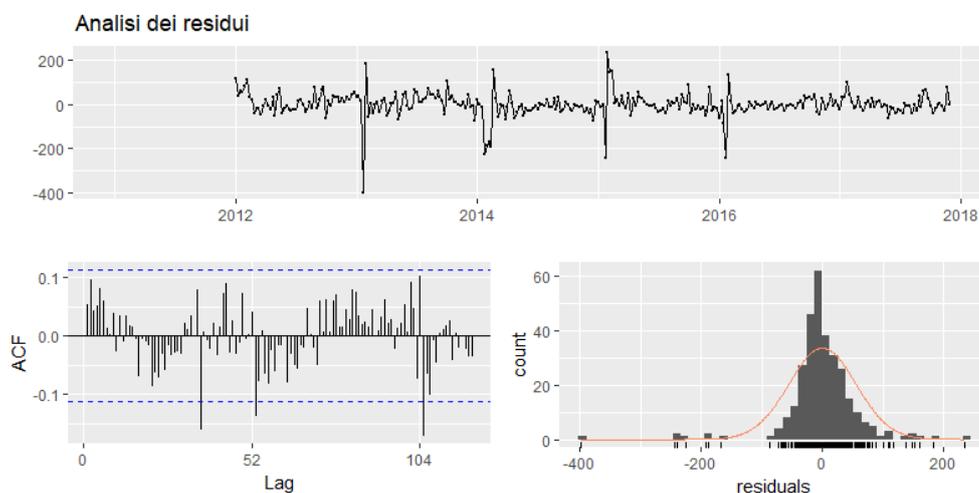


Figura 5.24

Vengono infine presentate le previsioni puntuali in pezzi di tutti i modelli presi in esame con l'evidenziazione del modello scelto che in questo caso è la regressione lineare. Inoltre l'ultima colonna evidenzia la previsione in colli delle settimane promozionali (7 e 8). Si nota come la previsione in colli per la regressione lineare non differisca molto dalla previsione del modello ARIMA, mentre è leggermente diversa dalla rete neurale.

Modelli	Settimane di previsione								Colli Promo
	1	2	3	4	5	6	7	8	
Regressione lineare	304	270	270	252	245	317	823	754	198
Regressione log lineare	277	254	250	238	231	275	759	709	184
ARIMA	291	278	275	258	236	262	797	746	193
Croston	266	266	266	266	266	266	809	809	203
Holt-Winters	316	297	287	266	268	325	946	312	158
Rete Neurale	285	280	284	266	233	239	739	695	180
CF	NA	NA	NA	NA	NA	NA	NA	NA	190

Tabella 5.12

Per completezza di informazione si riporta la figura 5.26 in cui si vede la serie storica originale sovrapposta alla serie stimata della regressione lineare, ovvero il modello scelto, rappresentata da una linea tratteggiata. Le promozioni sono evidenziate da dei pallini pieni mentre la previsione si trova alla fine della serie con gli ultimi 2 valori, ovvero le settimane 7 e 8.

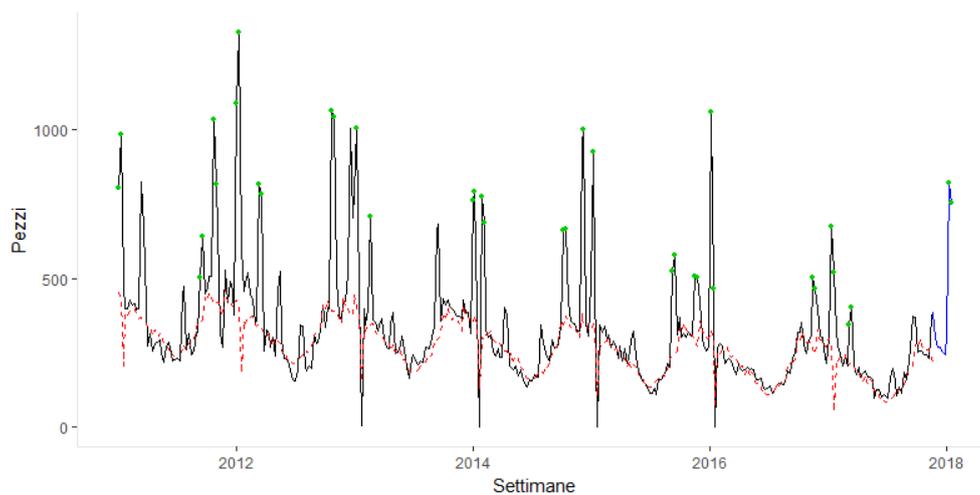


Figura 5.26

Anche in questo caso lo SKU è andato in promozione per 30 PDV, e a tutti i negozi viene assegnato almeno un collo. Tuttavia la previsione di 198 colli è una sovrastima in quanto, a consuntivo, sono stati venduti solo 150 colli.

Questo è dovuto a scelte di natura logistica, infatti il prodotto, essendo molo ingombrante, è stato consegnato ai PDV in alcune trance e non interamente nel primo giorno di effettiva promozione. Ciò ha comportato una rottura di stock in alcuni PDV per alcuni giorni della settimana promozionale e di conseguenza vendite più ridotte rispetto alla previsione proposta dai Demand Planner.

6. Conclusione

Come si è discusso in questo elaborato il ruolo del Demand Planner nei contesti aziendali della GDO assume una figura di importanza strategica in quanto riesce a dare un'indicazione di previsione di vendita dei singoli prodotti in tempi molto rapidi.

Si è mostrato come risolvere il problema delle previsioni promozionali grazie all'applicazione di tecniche di modellazione statistica standard ampliate con un metodo di stima dell'impatto promozionale, innovativo per la letteratura. Inoltre si è visto come utilizzare tecniche di valutazione dell'accuratezza previsionale specifiche per il problema in questione come lo "one-step-promotional-ahead forecast".

Ulteriori sviluppi di questa tesi possono riguardare l'implementazione di nuove tecniche di sales forecasting in aggiunta a quelle già mostrate; l'individuazione di nuove misure di accuratezza degli errori. Oppure l'ideazione di nuovi metodi di combinazione di previsioni provenienti da differenti modelli statistici.

Focus particolare è stato dato al fatto di non poter verificare chiaramente a consuntivo se le previsioni fornite dai Demand Planner siano state il più possibile corrette oppure no a causa di scelte manageriali o logistiche, che non sono direttamente controllabili dai Demand Planner. Ciò risulterebbe molto utile per la correzione dei modelli di sales forecasting o l'implementazione di nuovi, considerando ulteriori variabili come per esempio la giacenza dei punti vendita, oppure la rotazione delle scorte di magazzino.

Tuttavia bisogna tenere presente che in una realtà aziendale strutturata devono coesistere molti attori coinvolti nel processo previsionale e molto spesso questi non dialogano tra di loro a causa dell'appartenenza ad aree diverse dell'organigramma aziendale e ciò può essere molto limitante per la crescita dell'azienda stessa.

Riferimenti bibliografici:

- ❖ Dallari F., Creazza A., (2006), *Demand Planning. Risultati di un'indagine sul campo*. Logistica.
- ❖ Makridakis, S., Wheelwright, S. C., Hyndman, R. J, (1998), *Forecasting: Methods and Applications*, J. Wiley e Sons.
- ❖ Blatterg R C, Briesch R, and Fox, E J, (1995). *How Promotions Work*. Marketing Science 14: 122-132.
- ❖ Christen M, Gupta S, Porter J C, Staelin R, and Wittink D R, (1997). *Using-Market Level Data to Understand Promotion Effects in a Nonlinear Model*. Journal of Marketing Research 34: 322-334.
- ❖ Lattin J M, and Bucklin R E, (1989). *Reference Effects of Price and Promotion on Brand Choice Behavior*. Journal of Marketing Research 26: 299-310.
- ❖ Mulhern F J, and Leone R P, (1991). *Implicit Price Building of Retail Products: a Multiproduct Approach to Maximizing Store Profitability*. Journal of Marketing 55: 63-76.
- ❖ Cooper L G, and Giuffrida G, (2000). *Turning Data Mining into a Management Science Tool: New Algorithms and Empirical Results*. Management Science 46: 249-264.
- ❖ Baltas G (2005). *Modeling Category Demand in Retail Chains*. Journal of the Operational Research Society, 56:1258-1264.
- ❖ Narasimhan C, Neslin S A, and Sen S K, (1996). *Promotion Elasticities and Category Characteristics*. Journal of Marketing 60: 17-30.
- ❖ Struse R W, (1987). *Commentary: approaches to promotion evaluation: a practitioner's viewpoint*. Marketing Science 6:150-151.
- ❖ Neslin S A, (1990). *A market response model for coupon promotions*. Marketing Science 9: 125-145.
- ❖ Di Fonzo, T., Lisi, F., (2005), *Serie storiche ed economiche: analisi statistiche e applicazioni*. Roma: Carrocci editore.

- ❖ Wooldrige, J. M., (2006), *Introductory Econometrics: A Modern Approach*. 4° ed. Mason: South-western Cengage learning. pp. 339-376.
- ❖ Milanato, (2008), *Demand Planning: Processi, metodologie e modelli matematici per la gestione della domanda commerciale*.
- ❖ Hyndman, R. J., (2014), *Measuring forecast accuracy*. Disponibile su: <<http://www.robjhyndman.com/papers/forecast-accuracy.pdf>>
- ❖ Nau, R., (2014) *Forecasting with moving averages*. Disponibile su: https://people.duke.edu/rnau/Notes_on_forecasting_with_moving_averages-Robert_Nau.pdf.
- ❖ Toffalis, C., (2015), *A better measure of relative prediction accuracy for model selection and model estimation*. Journal of the Operational Research Society, Vol. 66,1352-1362.
- ❖ Armstrong, J.S., Collopy, F., (1992), *Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons*. International Journal of Forecasting, Vol. 8,69,80.
- ❖ Armstrong, J. S., (2008), *A Commentary on Error Measures*. International Journal of Forecasting Vol. 8. pp. 99-111.
- ❖ Trusov, M., Bodapati, A.V., Cooper, L.G., (2006), *Retailer Promotion Planning: Improving Forecast Accuracy and Interpretability*, Journal of Interactive Marketing, Vol.20, No.3-4,pp. 71-81.
- ❖ Brown, R. G., (1956), *Exponential smoothing for predicting demand*. ORSA, Nov.
- ❖ Holt, C. C., (1957), *Forecasting Trends and Seasonals by Exponentially Weighted Averages*, Carnegie Institute of Technology, Pittsburgh Office of Naval Research memorandum no. 52
- ❖ Winters, P. R., (1960), *Forecasting Sales by Exponentially Weighed Moving Averages*. Management Science. Volume 6. pag 324-342.
- ❖ Croston J.D., (1972), *Forecasting and stock control for intermittent demands*. Operational Research Quarterly. Volume 23. pag 289-304.

- ❖ Brockwell P. j., Davis R. A., (1991), *Time Series: Theory and Methods*, 2nd ed., Springer-Verlag, New York.
- ❖ Wei W. W. S., (1990), *Time series analysis. Univariate and multivariate methods*, Addison-Wesley, Redwood City.
- ❖ Box G. E. P., Jenkins G. M., (1976), *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco.
- ❖ Akaike H., (1974), *A new look at the statistical model identification*, "IEEE Transactions on Automatic Control", AC-19, pp. 716-723.
- ❖ Azzalini A., (1992), *Inferenza Statistica*, Springer-Verlag, Berlino.
- ❖ Schwarz G., (1978), *Estimating the dimension of a model*, "The Annals of Statistics", 6, pp. 461-464.
- ❖ Hamilton J. D., (1994), *Time Series Analysis*, Princeton University Press New Jersey. Ed. Italiana: *Econometria delle serie storiche*, cura di B. Sitzia, Monduzzi, Bologna, 1995.
- ❖ Siegel S. Castellan N. J., (1992), *Statistica non parametrica*, McGraw-Hill, Milano.
- ❖ Azzalini, A. e Scarpa, B., (2004). *Analisi dei dati e data mining*. Springer-Verlag, Milano.
- ❖ Venables, W. N. e Ripley, B. D., (1997). *Modern applied statistics with S-PLUS*. Springer-Verlag, II edition.
- ❖ Silver, E.A., Pyke D.F., Peterson R., (1998). *Inventory Management and Production Planning and Scheduling*. Third edition
- ❖ Chase W. C. Jr., (2013). *Demand-Driven Forecasting: a structured Approach to Forecasting*. Wiley.