



UNIVERSITA' DEGLI STUDI DI PADOVA

**DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI
"M. FANNO"**

DIPARTIMENTO DI SCIENZE STATISTICHE

CORSO DI LAUREA IN ECONOMIA

PROVA FINALE

**"Analisi di serie storiche e stima di andamenti futuri: modelli di riferimento
ed applicazioni nel campo dell'industria dell'occhialeria"**

RELATORE:

CH.MO/A PROF./SSA TOMMASO DI FONZO

LAUREANDO/A: ALESSIO ALFONSO

MATRICOLA N. 1216311

ANNO ACCADEMICO 2021 – 2022

Dichiaro di aver preso visione del “Regolamento antiplagio” approvato dal Consiglio del Dipartimento di Scienze Economiche e Aziendali e, consapevole delle conseguenze derivanti da dichiarazioni mendaci, dichiaro che il presente lavoro non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere. Dichiaro inoltre che tutte le fonti utilizzate per la realizzazione del presente lavoro, inclusi i materiali digitali, sono state correttamente citate nel corpo del testo e nella sezione ‘Riferimenti bibliografici’.

I hereby declare that I have read and understood the “Anti-plagiarism rules and regulations” approved by the Council of the Department of Economics and Management and I am aware of the consequences of making false statements. I declare that this piece of work has not been previously submitted – either fully or partially – for fulfilling the requirements of an academic degree, whether in Italy or abroad. Furthermore, I declare that the references used for this work – including the digital materials – have been appropriately cited and acknowledged in the text and in the section ‘References’.

Firma (signature)*Alessio Alfonso*.....

INDICE

INTRODUZIONE.....	4
CAPITOLO 1 - ANALISI DELLE SERIE STORICHE	5
1.1 - <i>Le serie storiche e le sue componenti.....</i>	5
1.2 - <i>Le tecniche di analisi</i>	10
CAPITOLO 2 - I METODI QUANTITATIVI DI BASE	11
2.1 - <i>Analisi dei residui.....</i>	13
2.2 - <i>Stime future.....</i>	14
2.3 - <i>Valutazione dell'accuratezza previsiva</i>	15
2.4 - <i>Valutazione della distribuzione delle stime</i>	16
CAPITOLO 3 - I MODELLI QUANTITATIVI COMPLESSI	18
3.1 - EXPONENTIAL SMOOTHING MODEL	18
3.1.1 - <i>La componente trend nel modello</i>	19
3.1.2 - <i>La componente stagionale nel modello.....</i>	20
3.1.3 - <i>Intervalli di stima con exponential smoothing models.....</i>	22
3.1.4 - <i>Selezione del modello.....</i>	24
3.1.5 - <i>Stima di andamenti futuri</i>	25
3.2 - AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODEL	27
3.2.1 <i>Moving Average models.....</i>	27
3.2.2 <i>Autoregressive models.....</i>	27
3.2.3 <i>Stazionarietà, differenziazione e random walk model</i>	28
3.2.4 <i>Invertibilità</i>	29
3.2.5 <i>ARMA model.....</i>	30
3.2.6 <i>ARIMA model.....</i>	30
3.2.7 <i>Seasonal ARIMA.....</i>	31
3.2.8 <i>Funzione di autocorrelazione e funzione di autocorrelazione parziale.....</i>	31
3.2.9 <i>Selezione del modello.....</i>	32
3.2.10 <i>Stime future.....</i>	32
CAPITOLO 4 - APPLICAZIONE DEI MODELLI.....	33
4.1 <i>APPLICAZIONE DELL'EXPONENTIAL SMOOTHING MODEL</i>	36
4.2 <i>APPLICAZIONE DEL MODELLO ARIMA</i>	40
4.3 <i>CONFRONTO SEASONAL NAÏVE, ETS E ARIMA</i>	43
CONCLUSIONI	44
BIBLIOGRAFIA	45

Introduzione

La tesi tratta il tema dell'analisi di serie storiche e della loro previsione futura. Questo lavoro ha come obiettivo l'applicazione di tali strumenti a dati di vendita del settore dell'occhialeria. La tesi è organizzata nel modo seguente. Nel primo capitolo, viene fornita una introduzione generale su che cosa si intende per analisi di serie storiche, i vari contesti di utilizzo e le loro principali caratteristiche. Nel secondo capitolo vengono descritte le tre tecniche di base per analizzare i dati passati con l'obiettivo di stimare andamenti futuri: il mean method, il naïve method e il drift method. Tali modelli verranno presi in considerazione come benchmark per giustificare l'utilizzo di modelli più complessi nella stima delle vendite future. Nel terzo capitolo si passa ai modelli complessi: il modello lisciamente esponenziale e il modello ARIMA verranno approfonditi secondo le loro caratteristiche principali. Nel quarto capitolo si applicano tali modelli ai dati forniti da Safilo spa attraverso l'utilizzo del software statistico R.

Ovviamente, il tema preso in considerazione è molto ampio, quindi una selezione dei modelli da prendere in considerazione per la loro spiegazione era d'obbligo. Il documento non ha come obiettivo quello di coprire tutto il tema, ma bensì di evidenziare le principali metodologie ed applicarle ad uno specifico contesto.

Capitolo 1 - Analisi delle serie storiche

1.1 - Le serie storiche e le sue componenti

Una serie storica viene definita come una sequenza cronologica di osservazioni di una determinata variabile di interesse (Montgomery, 2015). I valori della variabile possono essere osservati giornalmente, mensilmente, trimestralmente e così via a seconda della cadenza temporale con cui la variabile è stata osservata. Vari sono i contesti nei quali l'analisi delle serie storiche per la stima di possibili andamenti futuri è utilizzata.

Alcuni suoi esempi di utilizzo possono essere:

- Nell'operation management, dove si stima la domanda di servizi e/o la vendita di determinati prodotti in modo da programmare la produzione, l'approvvigionamento delle materie prime, il personale necessario ecc.;
- Nel marketing, per prevedere possibili risposte a campagne pubblicitarie;
- Nel mondo della finanza e del risk management, nei quali uno degli obiettivi potrebbe essere quello di stimare il tasso di redditività per determinati investimenti;
- In macroeconomia, dove le previsioni di variabili come PIL, inflazione e disoccupazione aiutano i policy makers nelle decisioni di politica economica;
- Nel controllo dei processi industriali per la stima di standard qualitativi di un determinato processo produttivo;
- Nel campo demografico per stimare andamenti di nascite, morti, movimenti migratori e così via.

Nell'analisi di serie storiche è importante definire alcune delle caratteristiche che si possono riscontrare in esse. Alcune di queste sono la presenza di trend, stagionalità e ciclicità.

Per trend si intende un incremento o decremento dei valori dei dati nel lungo periodo.

Nella figura 1.1 sono raffigurate le vendite di un brand di occhiali in una specifica area, che, per motivi di privacy, chiameremo Brand I dell'area X (i nomi fittizi verranno poi mantenuti per tutto il resto del documento). Le vendite sono divise per trimestre e vanno dal 2012 al 2019. In questi sette anni il grafico evidenzia come sia presente un netto trend decrescente nelle vendite.

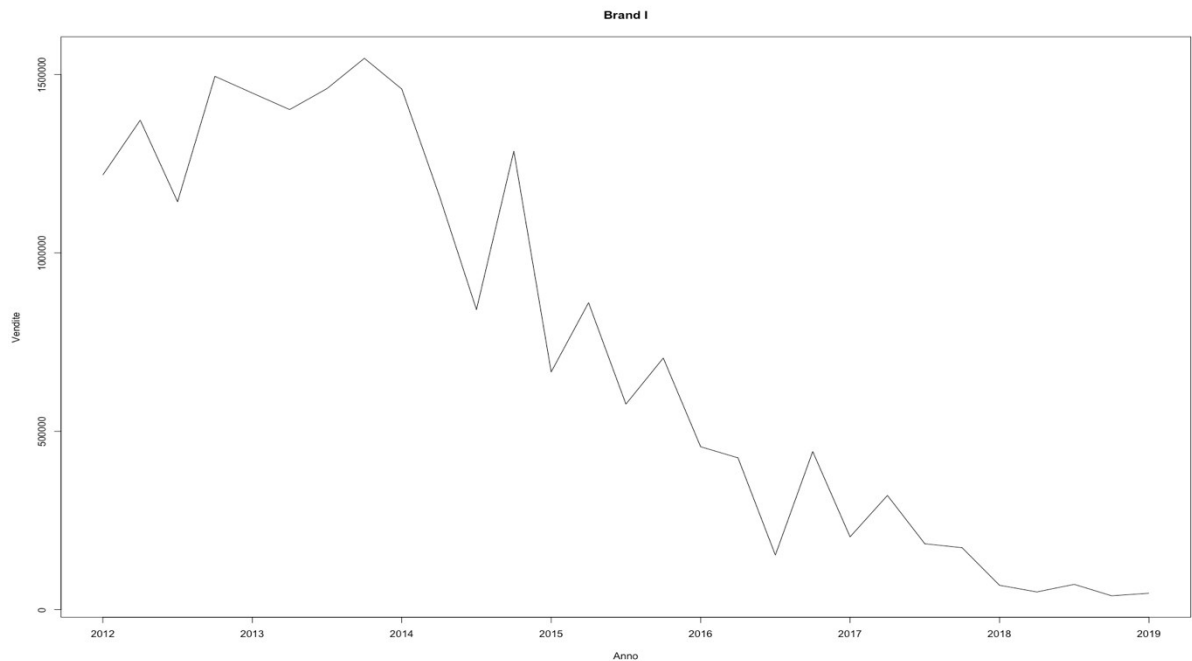


Figura 1.1: Vendite trimestrali Brand I – Area X

Per stagionalità si intende che la dinamica della serie storica si ripete in maniera abbastanza regolare a seconda del periodo in cui la serie viene osservata, ad esempio nello stesso giorno di settimane diverse o negli stessi mesi di anni diversi.

Nella figura 1.2 sono raffigurate le vendite di un altro brand, anch'esso per trimestre dal 2012 al 2019. In questo caso, oltre che un trend leggermente decrescente, è possibile notare come negli anni sia presente una stagionalità. In particolare, il secondo trimestre di ogni anno evidenzia vendite superiori rispetto a quelle degli altri trimestri.

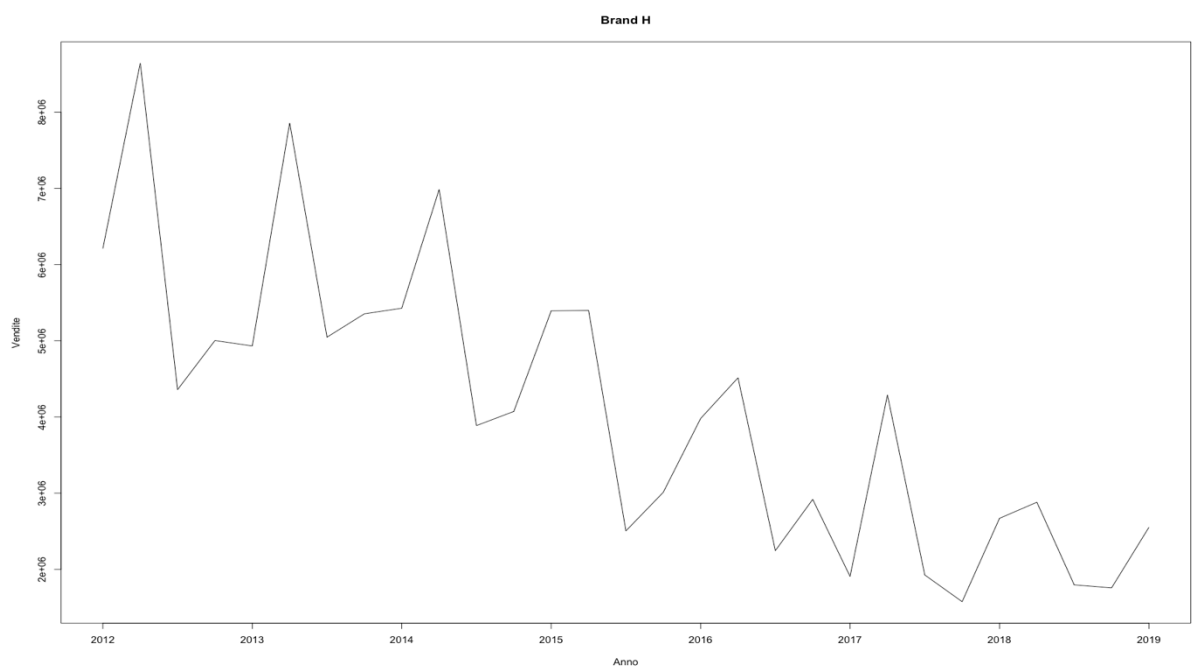


Figura 1.2: Vendite trimestrali Brand H – Area X

La stagionalità non va confusa con la componente ciclica. Quest'ultima è infatti identificabile in un andamento non prevedibile che si manifesta in aumenti e/o cali dei valori dei dati.

La ciclicità è sicuramente più visibile quando a disposizione si hanno serie storiche abbastanza lunghe. Nell'esempio riportato in figura 1.3, che riporta le vendite trimestrali del Brand D, è possibile notare come sul finire del 2014 sino alla metà del 2017 sia presente un ciclo al di sopra della media delle vendite rispetto all'intero periodo considerato.

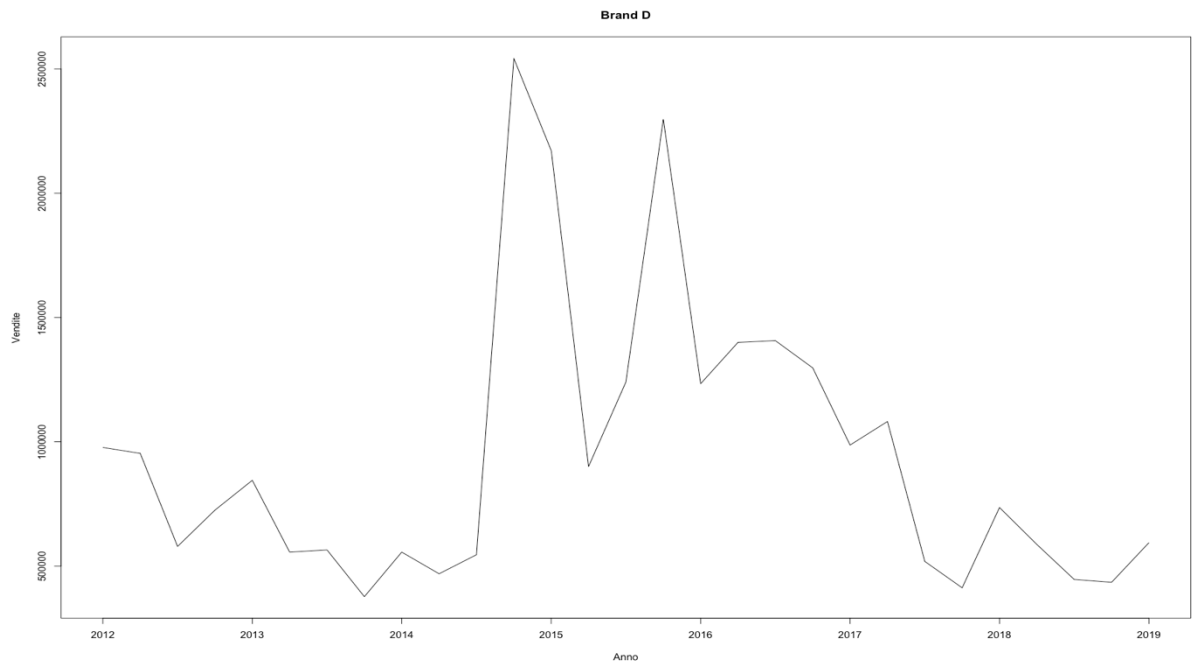


Figura 1.3: Vendite trimestrali Brand D – Area Q

Possiamo quindi decomporre la serie storica in modo da analizzare le sue componenti.

Definendo quindi la componente stagionale come S_t , la componente trend-ciclica come T_t e la componente residuale come R_t , possiamo considerare una decomposizione additiva $y_t = S_t + T_t + R_t$ (utilizzata nel caso in cui la componente stagionale non sia marcata e/o non ci sia molta variazione nella componente trend-ciclica), oppure moltiplicativa $y_t = S_t \times T_t \times R_t$ (spesso utilizzata per le serie economiche).

A questo punto occorre stimare le varie componenti della variabile considerata.

Il trend-ciclo può essere stimato utilizzando la Media Mobile, MA di ordine m , definita come:

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^k y_{t+j},$$

dove k è il numero di periodi di y_t ed m è dato da $2k + 1$. La stima del trend-ciclo è dunque una media dei valori dei dati della serie storica utilizzando k periodi. Questa tecnica permette

di livellare i dati delle serie storiche. Più elevato sarà il valore di m , più livellati saranno i dati ottenuti.

Una prosecuzione della MA è l'utilizzo di MA di MA. Una MA di ordine 2X4 è una MA di ordine 4 sulla quale viene ricalcolata successivamente una MA di ordine 2. L'utilità della doppia applicazione deriva dalla possibilità di livellare ulteriormente i dati oltre che alla ponderazione di essi in modo tale che i valori escano ed entrino nella formula in maniera graduale. Infatti, una MA di MA può essere definita come una Weighted MA:

$$\hat{T}_t = \sum_{j=-k}^k a_j y_{t+j},$$

dove a_j sono i pesi assegnati ai dati. Tali pesi sono simmetrici e loro somma deve essere uguale a 1.

Approfondita la modalità di stima del Trend, possiamo decomporre la serie storica utilizzando diverse metodologie.

La prima metodologia, ideata nel 1920 e base per altre metodologie sviluppatesi successivamente, è la decomposizione classica. In questa tipologia, una volta ottenuto \hat{T}_t , si può dunque calcolare $y_t - \hat{T}_t$. \hat{S}_t si assume costante di anno in anno ed il periodo di stagionalità viene definito come m , che sarà uguale a 4 nel caso di dati quadrimestrali, 12 per dati mensili, 7 per dati giornalieri riferiti ad una settimana completa, e così via. S_t si ottiene facendo una media dei valori precedenti dei dati della stagione considerata e poi questo dato viene replicato in sequenza per ogni anno della serie storica. L'ultimo step è quello di calcolare \hat{R}_t . Questa componente è calcolata come $\hat{R}_t = y_t - \hat{T}_t - \hat{S}_t$. Il procedimento appena elencato fa riferimento al caso di una decomposizione additiva, nel caso di una decomposizione moltiplicativa, i valori andranno divisi invece che sottratti.

Nella figura 1.4, sono state prese in considerazione le vendite trimestrali di un brand per identificarne le sue componenti. La decomposizione attuata è classica e di tipo moltiplicativo in quanto i dati considerati derivano da serie storiche con presenza di stagionalità.

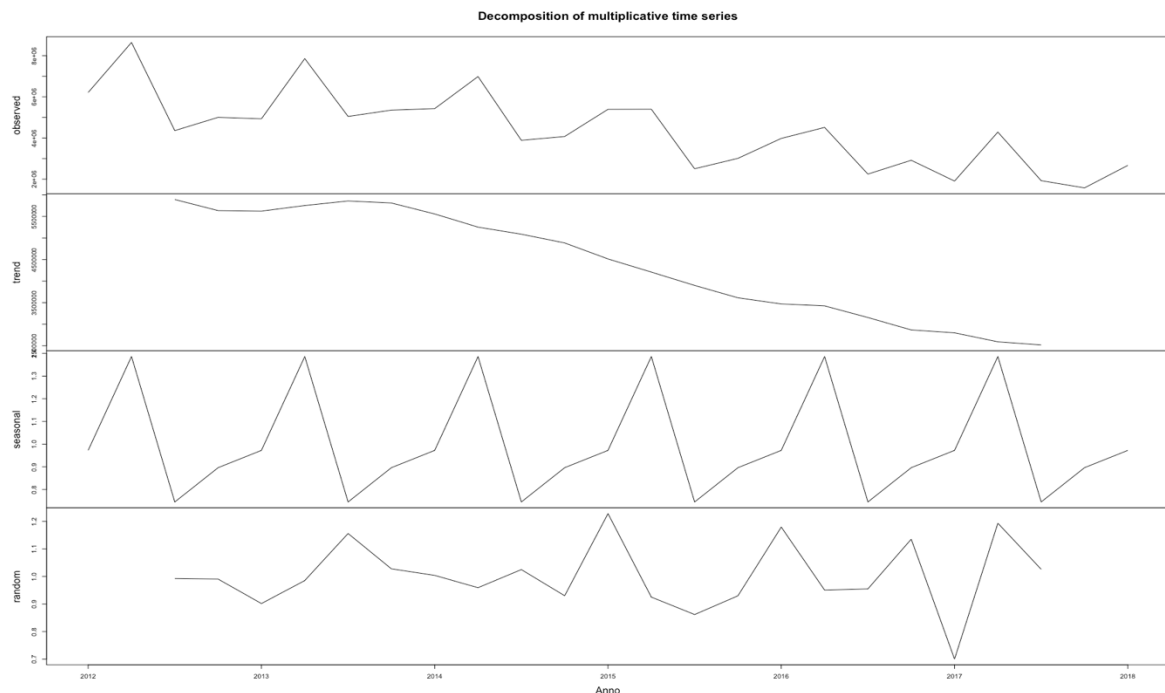


Figura 1.4: Decomposizione classica di tipo moltiplicativo Brand H – Area Y

Il primo grafico in alto mostra l'andamento di tali vendite, mentre gli altri tre mostrano rispettivamente trend, stagionalità e residui. La somma dei tre grafici sottostanti al primo danno come risultato il primo grafico.

La decomposizione classica non risulta però priva di problematiche, infatti:

- La stima della componente trend-ciclo non può essere effettuata per le prime e per le ultime osservazioni e, conseguentemente, anche per la componente residuale;
- La stima della componente trend-ciclica rischia di livellare troppo incrementi e decrementi repentini dei dati;
- La componente stagionale si assume ripetersi di anno in anno, ma per serie storiche molto lunghe ciò potrebbe non essere vero;
- Non è un metodo robusto per trattare valori anormali (outlier), cioè dati che si presentano nella serie storica in maniera non preventivata;

I metodi più utilizzati per stimare le componenti di serie dati mensili e trimestrali (Dagum, 2016) sono l'X-11 method e il SEATS method. Entrambi i modelli evitano quelli che sono i difetti della decomposizione classica, in particolare l'X-11 method effettua preventivamente

degli aggiustamenti sui dati (come, per esempio, le differenze tra giorni lavorativi mensili), ne calcola una MA e successivamente un'altra MA solitamente per 3 o 5 periodi. Una volta ottenuti i dati, identifica e sostituisce gli outliers con dati in linea con il Trend. Questo procedimento viene poi ripetuto nuovamente per ottenere dati ancor più livellati (Evans, 2002).

1.2 - Le tecniche di analisi

Le tecniche di analisi e previsione futura dei possibili andamenti di serie storiche possono essere suddivise in due macrocategorie: tecniche qualitative e tecniche quantitative.

Le tecniche qualitative sono soggettive e si basano sul giudizio di soggetti esperti del contesto al quale si applica la tecnica. Le tecniche qualitative si utilizzano in ambiti nei quali i dati passati sono assenti, ridotti o incompleti, oppure vengono usate in combinazione con le tecniche quantitative, sia per confrontare i risultati delle due tipologie, oppure per utilizzare in sequenza prima una tecnica quantitativa, per poi aggiustare l'output attraverso l'utilizzo di una tecnica qualitativa. Esempi possono essere il lancio di un nuovo prodotto o di una nuova tecnologia.

Le tecniche quantitative si basano invece sui dati storici e utilizzano modelli per analizzarli, identificare le caratteristiche e proiettarle per stimare l'andamento futuro della variabile considerata. I tre metodi base che fanno parte di questa categoria sono il moving average method, il naïve method e il drift method. I tre metodi precedenti più i modelli più complessi come l'exponential smoothing model e l'autoregressive integrated moving average model (ARIMA model) saranno spiegati nei capitoli successivi.

Capitolo 2 - I metodi quantitativi di base

I tre metodi quantitativi di base per l'analisi di dati e per le stime future sono il mean method, il naïve method e il drift method.

Il mean method effettua una stima futura basandosi sulla media dei valori passati. Andando quindi a definire i dati passati come y_1, y_2, \dots, y_T possiamo riscrivere la misurazione dei valori futuri come:

$$\hat{y}_{T+h|T} = \bar{y} = (y_1 + \dots + y_T) / T.$$

Utilizzando come esempio i dati del brand H visti in precedenza, una rappresentazione del mean method è raffigurata nella figura 2.1.

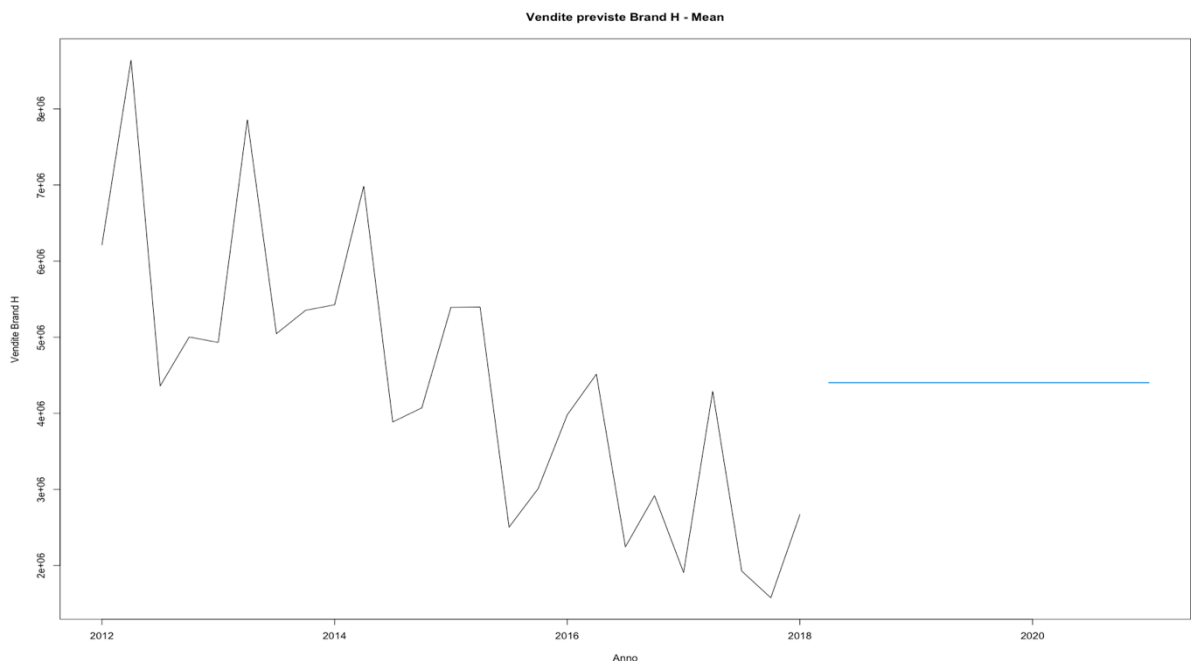


Figura 2.1: Vendite previste con mean method Brand H – Area Y

Il naïve method basa invece le stime future proiettando l'ultima osservazione ottenuta nel futuro (figura 2.2). Formalmente:

$$\hat{y}_{T+h|T} = y_T.$$

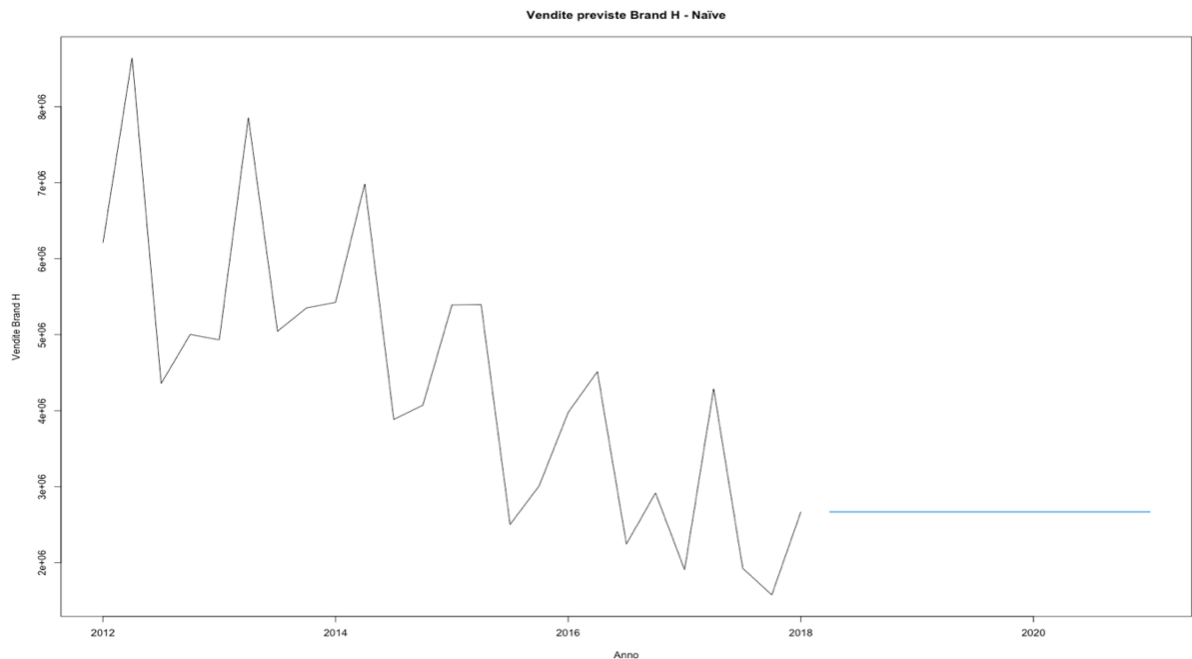


Figura 2.2: Vendite previste con naïve method Brand H – Area Y

Una variazione del naïve method è il seasonal naïve method (figura 2.3), che viene utilizzato in caso di stagionalità spiccata. In questo caso, ogni valore futuro prenderà lo stesso valore della serie storica nel periodo precedente ovvero nel caso in cui i dati siano mensili, il valore futuro di un determinato mese sarà uguale al valore dalla serie storica nello stesso mese dell'anno precedente. Il seasonal naïve method può essere definito come:

$$\hat{y}_{T+h|T} = y_{T+h-m}$$

dove m è il periodo di stagionalità.

Il drift method può essere considerato come un'altra variazione del naïve method, infatti, in

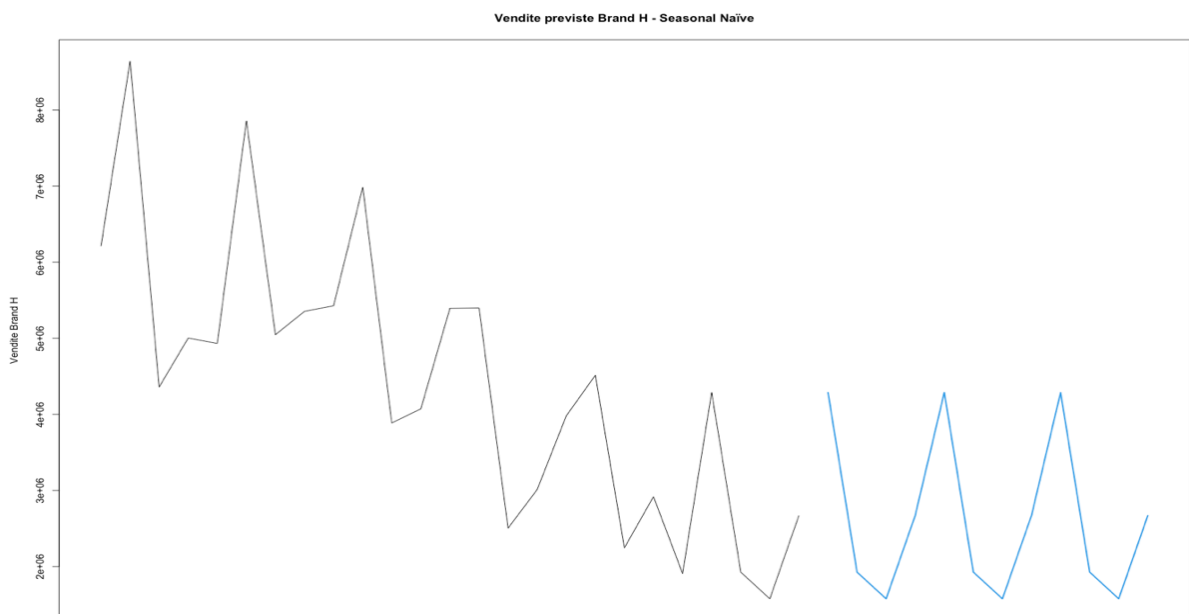


Figura 2.3: Vendite previste con seasonal naïve Brand H – Area Y

questo caso viene consentito alle stime future di cambiare il loro valore nel tempo dove il

cambiamento (drift) non è altro che la tendenza dei valori che si è osservata in passato proiettata nel futuro. Ciò è visibile nell'esempio riportato in figura 2.4: l'andamento delle vendite del Brand H mostra una diminuzione negli anni, il drift method presume quindi che questo calo continui anche in futuro, da qui la previsione in calo anche delle vendite future. Il concetto della stima dei valori futuri può anche essere interpretato come una retta passante tra il primo e l'ultimo valore della serie storica. La formula del drift method è definita come:

$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1} \sum_{t=2}^T (y_t - y_{t-1}) = y_T + h \left(\frac{y_T - y_1}{T-1} \right).$$

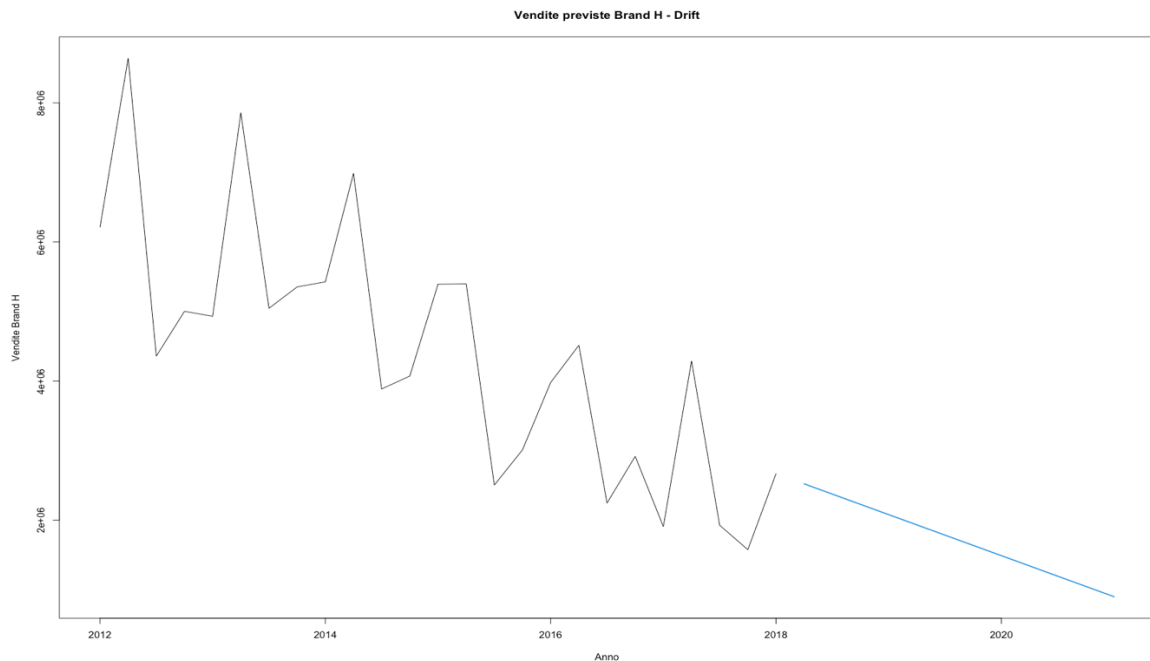


Figura 2.4: Vendite previste con drift method Brand H – Area Y

2.1 - Analisi dei residui

Per ognuno di questi modelli è opportuno analizzare i residui. I residui sono le differenze tra i valori effettivi dei dati e i valori stimati. Formalmente, i residui sono definiti come $e_t = y_t - \hat{y}_t$.

Affinché un modello possa essere considerato buono e/o migliorabile, è necessario che i residui rispettino due condizioni:

- Incorrelazione tra di essi (in caso di correlazione, infatti, significherebbe che informazioni utili sono state tralasciate dal nostro modello);
- Media 0 (in caso si media diversa da zero, il modello sarebbe distorto).

Nel caso in cui i residui presentino una media diversa da 0, una soluzione per migliorare il modello è quella di sottrarre il valore della media in tutte le osservazioni.

Per quanto riguarda la prima condizione, l'autocorrelazione dei residui è definita come:

$$\rho_k = \frac{\sum_{t=k+1}^T (e_t - \bar{e})(e_{t-k} - \bar{e})}{\sum_{t=1}^T (e_t - \bar{e})^2}$$

ed è possibile verificarne la presenza utilizzando la Autocorrelation Function (ACF), una funzione che mostra le autocorrelazioni per ogni ritardo, oppure attraverso i Portmanteau tests. Questi test permettono di capire se le prime ℓ autocorrelazioni sono significativamente diverse da quelle di un processo di white noise.

Un processo di white noise è una sequenza di variabili casuali con media 0, varianza costante uguale a σ^2 e autocovarianza (quindi anche autocorrelazione) uguale a 0 (M. Guidolin, 2018).

I Portmanteau test sono il Box – Pierce e il Ljung – Box.

Il Box – Pierce test è definito come:

$$Q = T \sum_{k=1}^{\ell} \rho_k^2 \sim \chi_{\ell-k}^2$$

dove T è il numero di osservazioni e ℓ è il massimo lag considerato. Sotto l'ipotesi nulla di assenza di autocorrelazione, il test si distribuisce come una v. c. Chi-Quadrato con $\ell-k$ gradi di libertà, dove l'ipotesi nulla è data da $H_0: \rho_1 = \rho_2 = \dots = \rho_{\ell} = 0$, mentre l'ipotesi alternativa è $H_1: \rho_{\ell} \neq 0$. Se il valore della statistica Q eccede il valore critico derivante dalla distribuzione Chi-Quadrato, allora l'ipotesi nulla viene rifiutata, quindi almeno una correlazione tra tutte quelle considerate è diversa da 0.

Una formula simile è applicata nel Ljung – Box test. Questo test è da preferirsi in caso di campioni più piccoli. La formula del test è

$$Q^* = T(T+2) \sum_{k=1}^{\ell} (T-k)^{-1} \rho_k^2 \sim \chi_{\ell-k}^2.$$

Anche in questo caso, valori elevati di Q^* portano a rispettare l'ipotesi che i residui siano significativamente diversi da un processo di white noise.

Altre due condizioni che possono facilitare l'analisi dei residui sono la presenza di omoschedasticità tra i residui e che essi si distribuiscano come una normale.

2.2 - Stime future

Possiamo determinare un intervallo di confidenza per i valori previsti.

Un intervallo di confidenza è definito come $\hat{y}_{t+h|T} \pm c \hat{\sigma}_h$ dove $\hat{\sigma}_h$ è la stima della deviazione standard nel periodo futuro $h+1$, calcolata come $\sqrt{\frac{1}{T-K} \sum_{t=1}^T e_t^2}$ e c è il valore critico della distribuzione normale corrispondente al livello di confidenza scelto.

Quando si vuole costruire un intervallo di confidenza per più anni (o mesi nel caso in cui la serie sia mensile, trimestri se è trimestrale ecc.), la formula precedente non sarà più

sufficiente, ma, assumendo incorrelazione, si possono derivare formule per ognuno dei tre metodi precedentemente visti, ovvero:

- Mean method: $\hat{\sigma}_h = \sigma\sqrt{1 + 1/T}$
- Naïve method: $\hat{\sigma}_h = \sigma\sqrt{h}$
- Seasonal naïve method: $\hat{\sigma}_h = \sigma\sqrt{k + 1}$
- Drift method: $\hat{\sigma}_h = \sigma\sqrt{h(1 + \frac{h}{t})}$

Quando le assunzioni di incorrelazione e varianza costante sono valide, ma i residui non si possono considerare distribuiti come una v. c. Normale, possiamo ovviare a ciò ipotizzando che i residui futuri si distribuiscano come i residui precedenti. Formalmente y_t è definita come $y_t = \widehat{y_{t|t-1}} + e_t$ dove $e_t = y_t - \widehat{y_{t|t-1}}$. Di conseguenza $y_{T+1} = \widehat{y_{T+1|T}} + e_{T+1}$.

2.3 - Valutazione dell'accuratezza previsiva

Nella valutazione delle stime effettuate, occorre differenziare i dati tra training e test data. I primi sono utilizzati come base per la realizzazione delle stime, mentre i secondi sono utilizzati per valutare l'accuratezza dei dati.

Gli errori delle stime calcolate sono quindi definiti come differenza tra i valori osservati (test data) e le stime prodotte utilizzando i training data, formalmente $e_{t+n} = y_{T+n} - \hat{y}_{T+n|T}$.

Le misure scale-dependent (ovvero la cui misura e i cui valori dipendono dalla grandezza in termini assoluti dei dati utilizzati) più utilizzate per misurare l'accuratezza degli errori sono:

- Mean square error (MSE): $mean(e_t^2)$
- Root mean square error (RMSE): \sqrt{MSE}
- Mean absolute error (MAE): $mean(|e_t|)$
- Median absolute error (MdAE): $median(|e_t|)$

L'RMSE è preferibile al MSE in quanto ha valori sulla stessa scala dei dati. Storicamente RMSE e MSE sono stati molto popolari nella valutazione degli errori, anche se tendono ad essere molto sensibili agli outliers rispetto a MAE e MdAE.

Definendo il percentage error come $p_t = 100e_t/y_t$, le misure basate sui percentage errors sono:

- Mean absolute percentage error (MAPE): $mean(|p_t|)$
- Median absolute percentage error (MdAPE): $median(|p_t|)$
- Root mean square percentage error (RMSPE): $\sqrt{mean(p_t^2)}$
- Root median square percentage error (RMdSPE): $\sqrt{median(p_t^2)}$

Gli svantaggi di queste misure sono che:

- se $y_t = 0$ i risultati sono indefiniti, mentre se $y_t \cong 0$ i risultati assumono valori molto elevati;
- queste misure tendono a penalizzare di più gli errori positivi rispetto a quelli negativi.

A fronte di queste considerazioni, spesso vengono utilizzati il symmetric mean absolute

percentage error (sMAPE) calcolato come $mean(\frac{200|y_t - \hat{y}_t|}{y_t + \hat{y}_t})$ oppure il symmetric median

absolute error (SMdAPE) calcolato come $median(\frac{200|y_t - \hat{y}_t|}{y_t + \hat{y}_t})$. In questo caso, infatti, valori

piccoli di y_t hanno meno effetti su queste due misure, le quali non sono però esenti da difetti:

nel caso in cui y_t si avvicini a zero, anche \hat{y}_t si avvicinerà presumibilmente a zero, quindi il risultato finale sarà comunque influenzato dall'instabilità del calcolo.

Un'alternativa ai percentage errors, quando si confrontano serie storiche con modelli con unità differenti, è quella di usare gli scaled errors measures.

- Per una serie storica priva di stagionalità con stime prodotte con il naïve method l'errore privo di unità di misura è definito come $q_j = \frac{e_j}{\frac{1}{T-1} \sum_{t=2}^T |y_t - y_{t-1}|}$. Se q_j è inferiore ad 1, allora deriva da una stima migliore rispetto alla stima di un anno successivo effettuata con il naïve method, viceversa nel caso sia maggiore ad uno.;
- Nel caso in cui la stagionalità sia presente nella serie storica e si utilizzi il naïve method, lo scaled error è definito come $q_j = \frac{e_j}{\frac{1}{T-m} \sum_{t=m+1}^T |y_t - y_{t-m}|}$.

In entrambi questi casi, le misure da utilizzare ottenuti gli scaled errors sono:

- Mean absolute scaled error (MASE): $mean(|q_t|)$
- Median absolute scaled error (MdASE): $median(|q_t|)$
- Root mean squared scaled error (RMSSE): $\sqrt{mean(q_t^2)}$

Di queste misure il MASE è preferibile al RMSSE perché meno sensibile agli outliers e più facilmente interpretabile, ed al MdASE in quanto è meno variabile con pochi dati (Hyndman, 2006).

2.4 - Valutazione della distribuzione delle stime

Per la valutazione della bontà della distribuzione delle stime si possono usare i seguenti metodi:

- Quantile scores: con questo punteggio, andremo a calcolare l'accuratezza dell'intervallo di confidenza sui possibili dati futuri, in particolare andremo a stimare

nel tempo t i quantili con una probabilità p . Questo valore è genericamente definito come $f_{p,t}$. La formula per il calcolo del quantile score è dunque:

$$Q_{p,t} = \begin{cases} 2(1-p)(f_{p,t} - y_t), & \text{se } y_t < f_{p,t} \\ 2p(y_t - f_{p,t}), & \text{se } y_t > f_{p,t} \end{cases}$$

Un valore basso di $Q_{p,t}$ indica una miglior stima del quantile. A seconda poi di che percentuale si scelga nella definizione del quantile, il funzionamento del Quantile score tende a penalizzare o depenalizzare un certo tipo di valori, più precisamente quando $p > 0,5$, il quantile score darà maggior penalità a valori di y_t più grandi di $f_{p,t}$, viceversa per valori di p inferiori a 0,5;

- Continuous ranked probability score (CRPS): questa misura permette di evidenziare i quantile scores con probabilità p per tutta la distribuzione di stime;
- Winkler score: è un calcolo fatto sugli intervalli di confidenza dove se il valore di y_t cade al di fuori dell'intervallo, il winkler score avrà un punteggio pari alla lunghezza dell'intervallo più una penalità proporzionale a quanto distante sarà il valore trovato rispetto all'intervallo. La formula per il suo calcolo è

$$W_{\alpha,t} = \begin{cases} (u_{\alpha,t} - \ell_{\alpha,t}) + \frac{2}{\alpha}(\ell_{\alpha,t} - y_t), & \text{se } y_t < \ell_{\alpha,t} \\ (u_{\alpha,t} - \ell_{\alpha,t}), & \text{se } \ell_{\alpha,t} \leq y_t \leq u_{\alpha,t} \\ (u_{\alpha,t} - \ell_{\alpha,t}) + \frac{2}{\alpha}(y_t - u_{\alpha,t}), & \text{se } y_t > u_{\alpha,t} \end{cases}$$

Dato che le estremità degli intervalli di confidenza sono costruiti come $\ell_{\alpha,t} = f_{\alpha/2,t}$ e $u_{1-\alpha/2,t} = f_{1-\alpha/2,t}$, aggiungendo i valori dei quantili e dividendo per α si può ottenere il winkler score: $W_{\alpha,t} = (Q_{\alpha/2,t} + Q_{1-\alpha/2,t})/\alpha$;

- Comparazione utilizzando gli skill scores: in questo caso si prendono in considerazione i CRPS di vari modelli per capire quale modello permetta un maggior adattamento nelle stime rispetto agli altri. Utilizzando per esempio il naïve method e il drift method, la formula sarà la seguente:

$$\frac{CRPS_{naïve} - CRPS_{drift}}{CRPS_{naïve}}$$

Questo calcolo evidenzia quanto (e se) il drift method si adatti e riesca a prevedere meglio gli andamenti futuri rispetto al naïve method (Hyndman, 2021).

Capitolo 3 - I modelli quantitativi complessi

3.1 - Exponential smoothing model

L'exponential smoothing model si basa sull'idea di stimare i dati futuri della variabile di interesse, associando "pesi" decrescenti ai valori dati passati in maniera esponenziale, ovvero i dati più recenti avranno maggiore importanza e quindi un peso maggiore rispetto a quelli meno recenti.

Il simple exponential smoothing model infatti è definito come

$$\hat{y}_{t+1|t} = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \dots$$

dove α è un numero compreso tra 0 e 1 ed $\hat{y}_{t+1|t}$ è una media ponderata (decrescente) delle osservazioni di y passate. Il parametro α controlla il peso delle passate osservazioni nella stima dei valori futuri. Più alti saranno i valori di α , meno rilevanti saranno le osservazioni più distanti nel tempo. Il caso estremo in cui il valore di α sia uguale a 1 corrisponde al naïve method.

Una variante del modello appena descritto corrisponde al caso in cui $\hat{y}_{t+1|t}$ è dato da

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha)\hat{y}_{t|t-1}$$

ovvero una media ponderata dell'osservazione più recente di y e del valore della stima di y al tempo -1 . Il processo di calcolo si origina stimando $\hat{y}_{t|t-1}$ ($=\ell_0$ nell'equazione) e proseguendo per gli anni successivi. Formalmente:

$$\hat{y}_{2|1} = \alpha y_1 + (1 - \alpha)\ell_0$$

$$\hat{y}_{3|2} = \alpha y_2 + (1 - \alpha)\hat{y}_{2|1}$$

$$\hat{y}_{4|3} = \alpha y_3 + (1 - \alpha)\hat{y}_{3|2}$$



$$\hat{y}_{T|T-1} = \alpha y_{T-1} + (1 - \alpha)\hat{y}_{T-1|T-2}$$

$$\hat{y}_{T+1|T} = \alpha y_T + (1 - \alpha)\hat{y}_{T|T-1}$$

Sostituendo le equazioni corrispondenti nella parte destra della formula si ottiene:

$$\hat{y}_{3|2} = \alpha y_2 + (1 - \alpha)[\alpha y_1 + (1 - \alpha)\ell_0]$$

$$= \alpha y_2 + \alpha(1 - \alpha)y_1 + (1 - \alpha)^2 \ell_0$$

$$\hat{y}_{4|3} = \alpha y_3 + (1 - \alpha)[\alpha y_2 + \alpha(1 - \alpha)y_1 + (1 - \alpha)^2 \ell_0]$$

$$= \alpha y_3 + \alpha(1 - \alpha)y_2 + \alpha(1 - \alpha)^2 y_1 + \alpha(1 - \alpha)^3 \ell_0$$



$$\hat{y}_{T+1|T} = \sum_{j=0}^{T-1} \alpha(1 - \alpha)^j y_{T-j} + (1 - \alpha)^T \ell_0$$

e dato che per numeri elevati di T l'ultimo termine diventa sempre più piccolo, questa equazione riconduce a quella del simple exponential smoothing model.

Il simple exponential smoothing model può anche essere rappresentato attraverso le sue componenti. Le componenti da considerare sono il livello, definito come ℓ_t , il trend b_t e la componente stagionale s_t . Inserendo per semplicità solamente il livello, il modello diventa:

$$\hat{y}_{t+h|t} = \ell_t$$

$$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$$

dove la prima equazione è l'equazione di stima, mentre la seconda equazione rappresenta l'equazione del livello.

Considerando che per ogni modello α e ℓ devono essere scelti, la scelta di essi può essere effettuata minimizzando la somma degli errori al quadrato che deriva dal calcolo:

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

3.1.1 - La componente trend nel modello

Il simple exponential smoothing model può essere riscritto con l'inserimento di un'equazione che definisca il trend.

La sua formula, nota come Holt's linear trend method (Holt, 2004), è:

$$\hat{y}_{t+h|t} = \ell_t + hb_t$$

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$$

dove ℓ_t è la stima del livello della serie al tempo t, b_t è la stima della componente trend della serie al tempo t, α è il parametro lisciante del livello ed è compreso tra 0 e 1, mentre β^* è il parametro lisciante per il trend ed anch'esso è compreso tra 0 ed 1.

In questo caso le equazioni presenti sono tre ovvero l'equazione di stima, l'equazione per il livello, la quale rappresenta una media ponderata tra osservazioni di y_t e le stime del livello e del trend al tempo t-1, e l'equazione del trend che rappresenta una media ponderata della stima del trend al tempo t (dato dalla differenza tra ℓ_t e ℓ_{t-1}) e della stima del trend al tempo t-1.

Il limite del metodo appena illustrato sta nell'andamento del trend, infatti, in quel caso il trend avrà una costante crescita/decrecita nel futuro. Per smorzare questo effetto, un altro metodo è stato formulato:

$$\hat{y}_{t+h|t} = \ell_t + (\phi + \phi^2 + \dots + \phi^h) b_t$$

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$$

dove ϕ è il parametro che smorza l'effetto del trend ed è compreso tra 0 ed 1. Se $\phi = 1$ il modello sarà uguale a quello di Holt's, nel caso in cui ϕ sia compreso tra 0 ed 1, allora le stime nel futuro convergeranno a $\ell_T + \phi b_T / (1 - \phi)$ quando $h \Rightarrow \infty$ causando quindi la presenza di trend nei primi periodi, mentre nel futuro il trend verrà smorzato e i valori si assesteranno.

3.1.2 - La componente stagionale nel modello

L'aggiunta della componente stagionale al modello coincide con l'inserimento di un'altra equazione oltre a quella di stima, del livello e del trend.

Holt e Winter (Winters, 1960) hanno attribuito il parametro lisciante γ per la componente stagionale s_t .

Anche in questo caso, il metodo può essere calcolato secondo la modalità additiva, consigliata nel caso in cui la stagionalità si verifichi in maniera fortemente costante, oppure secondo la modalità moltiplicativa, caso preferibile nel momento in cui la stagionalità cambi proporzionalmente al cambiare del livello del modello. Nel primo caso la stagionalità sarà evidenziata in termini assoluti e sottratta nell'equazione del livello, mentre nel secondo caso la stagionalità sarà espressa in termini relativi e dunque divisa nell'equazione del livello.

Il modello additivo di Holt e Winter si presenta dunque così:

$$\hat{y}_{t+h|t} = \ell_t + h b_t + s_{t+h-m(k+1)}$$

$$\ell_t = \alpha(y_t - s_{t-m}) + (1 + \alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$$

$$s_t = \gamma(y_t - \ell_t - \ell_{t-1}) + (1 - \gamma)s_{t-m}$$

dove s_t è l'equazione per la componente stagionale, γ è il parametro lisciante della stagionalità ed m indica il periodo di stagionalità, il quale per esempio sarà uguale a 4, nel caso in cui i dati vengano divisi in trimestri, a 12 per i dati mensili e così via.

Spesso l'equazione della componente stagionale viene espressa come:

$$s_t = \gamma^*(y_t - \ell_t) + (1 - \gamma^*)s_{t-m}$$

ma se sostituiamo ℓ_t a questa equazione, si ottiene:

$$s_t = \gamma^*(1 - \alpha)(y_t - \ell_{t-1} - b_{t-1}) + [1 - \gamma^*(1 - \alpha)]s_{t-m}$$

che è uguale all'equazione specificata inizialmente per la componente stagionale con

$\gamma = \gamma^*(1 - \alpha)$. Il parametro γ^* dovrà quindi essere compreso tra 0 e 1, il che corrisponde per il parametro γ ad essere compreso tra 0 e $1 - \alpha$.

Analogamente il modello moltiplicativo di Holt e Winter sarà:

$$\begin{aligned}\hat{y}_{t+h|t} &= (\ell_t + hb_t)s_{t+h-m(k+1)} \\ \ell_t &= \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1} \\ s_t &= \gamma \frac{y_t}{(\ell_t + \ell_{t-1})} + (1 - \gamma)s_{t-m}\end{aligned}$$

Anche per questi due modelli, l'andamento del trend presenta lo stesso limite visto in precedenza con il metodo di Holt. Si può dunque ricorrere al parametro ϕ per smorzare gli effetti di tale formulazione. Una delle configurazioni più utilizzate è quella di smorzare l'effetto del trend utilizzando il parametro ϕ in un modello moltiplicativo. La formulazione sarà quindi la seguente:

$$\begin{aligned}\hat{y}_{t+h|t} &= [\ell_t + (\phi + \phi^2 + \dots + \phi^h)b_t]s_{t+h-m(k+1)} \\ \ell_t &= \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1}) \\ b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1} \\ s_t &= \gamma \frac{y_t}{(\ell_{t-1} + \phi b_{t-1})} + (1 - \gamma)s_{t-m}\end{aligned}$$

Una raffigurazione completa di tutte le possibili combinazioni tra l'inserimento del trend e l'inserimento della stagionalità nei modelli con il metodo additivo o moltiplicativo è raffigurata nella figura 3.1.1:

Trend Component	Seasonal Component		
	N	A	M
	(None)	(Additive)	(Multiplicative)
N (None)	(N,N)	(N,A)	(N,M)
A (Additive)	(A,N)	(A,A)	(A,M)
A_d (Additive damped)	(A_d ,N)	(A_d ,A)	(A_d ,M)

Figura 3.0.1.1: Combinazioni delle componenti modello ETS

I modelli associati a quelle combinazioni saranno quindi:

Trend	Seasonal		
	N	A	M
N	$\hat{y}_{t+h t} = \ell_t$	$\hat{y}_{t+h t} = \ell_t + s_{t+h-m(k+1)}$	$\hat{y}_{t+h t} = \ell_t s_{t+h-m(k+1)}$
	$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)\ell_{t-1}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)\ell_{t-1}$
		$s_t = \gamma(y_t - \ell_{t-1}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma(y_t/\ell_{t-1}) + (1 - \gamma)s_{t-m}$
A	$\hat{y}_{t+h t} = \ell_t + hb_t$	$\hat{y}_{t+h t} = \ell_t + hb_t + s_{t+h-m(k+1)}$	$\hat{y}_{t+h t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$
	$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$
	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$
Ad	$\hat{y}_{t+h t} = \ell_t + \phi_h b_t$	$\hat{y}_{t+h t} = \ell_t + \phi_h b_t + s_{t+h-m(k+1)}$	$\hat{y}_{t+h t} = (\ell_t + \phi_h b_t)s_{t+h-m(k+1)}$
	$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$
	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$
		$s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma(y_t/(\ell_{t-1} + \phi b_{t-1})) + (1 - \gamma)s_{t-m}$

Figura 3.1.2: Modelli ETS

3.1.3 - Intervalli di stima con modelli di lisciamento esponenziale

I modelli visti nel capitolo precedente permettono di generare stime puntuali. I modelli che si analizzeranno in questo capitolo permettono di generare intere distribuzioni di stima.

Ogni modello consiste in una equazione di misurazione e più equazioni di stato che analizzano i cambiamenti nel tempo delle componenti livello, trend e stagionalità. Da qui il nome di state space models.

Ogni modello contiene degli errori che possono essere calcolati con il metodo additivo o con il metodo moltiplicativo. I modelli in questione vengono denotati con tre lettere ovvero ETS (...,...,...) dove la prima lettera indica la modalità di calcolo degli errori, la seconda l'inserimento e la modalità di calcolo del trend, mentre la terza si riferisce alla stagionalità.

Il simple exponential smoothing model, per la previsione di intervalli di stima con l'inserimento degli errori calcolati con il metodo additivo, sarà definito come un ETS (A, N, N) dove la sua equazione di stime è data da:

$$\hat{y}_{t+1|t} = \ell_t$$

l'equazione di stato, in questo caso per definire il livello è invece data da:

$$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$$

dove, con i dovuti calcoli possiamo evidenziare il termine d'errore, ovvero:

$$\begin{aligned} \ell_t &= \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \\ &= \ell_{t-1} + \alpha e_t \end{aligned}$$

dove e_t è il residuo al tempo t ed è dato da $y_t - \ell_{t-1}$ ovvero $y_t - \hat{y}_{t|t-1}$.

Gli errori danno un'informazione importante in merito alle stime fatte nei periodi precedenti, infatti, se e_t è negativo, significa che $y_t < \hat{y}_{t|t-1}$ e che quindi il livello al tempo $t-1$ è stato

sovrastimato. Di conseguenza, il livello al tempo t dipende da α : più α sarà vicino ad 1, più il livello al tempo t sarà aggiustato, viceversa nel caso di α piccoli.

Un modello statistico completamente specificato si ottiene andando a definire la distribuzione degli errori. Riformulando l'equazione che definisce il termine d'errore, possiamo trovare che:

$$y_t = \ell_{t-1} + e_t$$

ovvero che y_t è data dal livello precedente più un errore. Assumendo che i residui (ovvero gli errori stimati per i periodi successivi) si distribuiscano come una normale, con media 0 e varianza σ^2 , possiamo riscrivere il modello come:

$$y_t = \ell_{t-1} + \varepsilon_t$$

$$\ell_t = \ell_{t-1} + \alpha \varepsilon_t.$$

L'equazione di misurazione mostra una relazione lineare tra la parte prevedibile di y , ovvero il suo livello precedente, e la parte imprevedibile. In questo caso si suppone che la relazione sia lineare, ma relazioni di tipo non lineare possono essere rappresentate dal modello.

L'equazione di stato rappresenta l'evoluzione nel tempo del livello che condiziona poi il valore di y . α gioca un ruolo fondamentale nel capire l'andamento del livello (e quindi di y).

Valori elevati di α daranno luogo a rapidi cambiamenti, mentre valori bassi tenderanno a stabilizzare l'andamento dei dati. Quando $\alpha = 0$, l'andamento non avrà cambi, mentre al contrario quando $\alpha = 1$ si è in presenza di random walk.

Nel caso in cui gli errori vengano calcolati con il metodo moltiplicativo, i residui saranno calcolati con la seguente formula:

$$\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}}$$

dove ε_t si distribuisce come una normale, con media 0 e varianza σ^2 . Se ad $\hat{y}_{t|t-1}$ sostituiamo

ℓ_{t-1} otteniamo $y_t = \ell_{t-1} + \ell_{t-1}\varepsilon_t$ e $e_t = y_t - \hat{y}_{t|t-1} = \ell_{t-1}\varepsilon_t$.

Il modello statistico completamente specificato sarà dunque:

$$y_t = \ell_{t-1}(1 + \varepsilon_t)$$

$$\ell_t = \ell_{t-1}(1 + \alpha \varepsilon_t).$$

Una panoramica completa di tutte le possibili combinazioni ottenibili, considerando anche il termine d'errore, è raffigurata nella figura 3.1.3:

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = \ell_{t-1} s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / \ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + b_{t-1})$
A _d	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = \phi b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + \phi b_{t-1})$

MULTIPLICATIVE ERROR MODELS

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha \varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1} s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha \varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma \varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha \varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha \varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma \varepsilon_t)$
A _d	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha \varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha \varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma \varepsilon_t)$

Figura 3.1.3: Modelli ETS con termine d'errore

Tra le varie combinazioni, nella tabella che considera gli errori con il metodo additivo, possiamo notare che la presenza del trend calcolata con il metodo additivo riconduce al metodo lineare di Holt. In questo caso i residui sono calcolati come: $\varepsilon_t = y_t - \ell_{t-1} - b_{t-1} \sim N(0, \sigma^2)$. La sua sostituzione nell'equazione originale darà luogo alla formulazione presente nella tabella dove $\beta = \alpha\beta^*$. Gli stessi concetti valgono anche per il metodo di Holt calcolato con errori moltiplicativi, dove l'unica differenza è appunto il calcolo dei residui ottenuto come segue: $\varepsilon_t = \frac{y_t - \ell_{t-1} + b_{t-1}}{(\ell_{t-1} + b_{t-1})} \sim N(0, \sigma^2)$.

3.1.4 - Selezione del modello

Per la scelta tra modelli occorre prima introdurre il concetto di massima probabilità, dato che figura in alcuni dei criteri di scelta. Questo concetto rappresenta un'alternativa alla scelta dei parametri in base alla minimizzazione della somma degli errori. Esso fonda la scelta dei parametri in tenendo conto di coloro che massimizzano maggiormente la probabilità del modello, ovvero alla probabilità di ottenerne uno che possa rappresentare al meglio i dati stimati. Per i modelli che utilizzano il metodo additivo per il calcolo degli errori, la suddetta

metodologia conduce agli stessi risultati della minimizzazione della somma degli errori, mentre per i modelli moltiplicativi ciò potrebbe non verificarsi.

I valori che i parametri α , β^* , γ^* , ϕ possono prendere sono ristretti. Essi devono essere compresi tra 0 e 1. Per i modelli state space con i parametri β e γ rispettivamente pari a $\alpha\beta^*$ e $(1-\alpha)\gamma^*$, le restrizioni saranno pari a: $0 < \alpha < 1$, $0 < \beta < \alpha$, $0 < \gamma < 1 - \alpha$.

Per quanto riguarda il parametro smorzante ϕ spesso il suo valore viene compreso tra numerosi software statistici tra 0,8 e 0,98. Un ulteriore approfondimento riguardo la scelta dei parametri si ottiene considerando le proprietà matematiche dei modelli state space. Infatti, l'obiettivo di dare meno peso ai dati più lontani nel tempo consente di stabilire delle nuove regioni di ammissibilità meno restrittive. Per esempio, considerando il modello ETS (A,N,N) dove il parametro α è normalmente compreso tra 0 e 1, la sua regione di ammissibilità sarà invece compresa tra 0 e 2 (Hyndman, 2008).

La selezione del modello avviene secondo tre criteri:

- Akaike's Information Criterion, calcolato come:

$$AIC = -2 \log(L) + 2k$$

dove L è la probabilità del modello, mentre k è il numero totale di parametri e stati/componenti inseriti nel modello per la stima comprensivi della varianza dei residui;

- Akaike's Information Criterion corretto, calcolato come:

$$AIC_c = AIC + \frac{2k(k+1)}{T-k-1}$$

- Bayesian Information Criterion, definito dalla formula:

$$BIC = AIC + k[\log(T) - 2]$$

Per ognuno di questi criteri, il modello migliore sarà quello con il valore minore. È da specificare che alcuni modelli che utilizzano il metodo moltiplicativo potrebbero causare instabilità nei calcoli quando si evidenzia la presenza di dati non strettamente positivi nella serie. Più precisamente, i modelli in questione sono ETS(A,N,M), ETS(A,A,M), e ETS(A,A_d,M).

3.1.5 - Stima di andamenti futuri

L'andamento dei dati futuri può essere ottenuto impostando semplicemente il tempo dell'equazione nei periodi successivi, ovvero $t = T+1, \dots, T+h$, e settando i residui dei periodi futuri uguali a 0 ($\varepsilon_t = 0$). L'azzeramento dei residui per i periodi successivi fa sì che determinati modelli producano le stesse previsioni. È il caso, per esempio, del modello di

Holt, il quale produce le stesse stime dei modelli ETS(A,A,N) e ETS(M,A,N). Prendendo quest'ultimo come esempio, la sua equazione di stima sarà:

$$y_{T+1} = (\ell_T + b_T)(1 + \varepsilon_{T+1})$$

la quale diventerà $\hat{y}_{T+1|T} = \ell_T + b_T$.

Per il periodo T+2 di conseguenza:

$$\begin{aligned} y_{T+2} &= (\ell_{T+1} + b_{T+1})(1 + \varepsilon_{T+2}) \\ &= [(\ell_T + b_T)(1 + \alpha\varepsilon_{T+1}) + b_T + \beta(\ell_T + b_T)\varepsilon_{T+1}](1 + \varepsilon_{T+2}) \end{aligned}$$

dove, con i dovuti calcoli, si ottiene $\hat{y}_{T+2|T} = \ell_T + 2b_T$.

Gli intervalli di previsione sono invece calcolati secondo la seguente formula:

$$\hat{y}_{T+h|T} \pm c\sigma_h$$

dove c è il valore critico individuato dalla distribuzione normale secondo una determinata probabilità, mentre σ_h è la varianza della stima. Per quest'ultima sono presenti delle formule rappresentate in figura 3.1.4 (non sono presenti formule per tutti i modelli).

Model	Forecast variance: σ_h^2
(A,N,N)	$\sigma_h^2 = \sigma^2 [1 + \alpha^2(h-1)]$
(A,A,N)	$\sigma_h^2 = \sigma^2 \left[1 + (h-1) \left\{ \alpha^2 + \alpha\beta h + \frac{1}{6}\beta^2 h(2h-1) \right\} \right]$
(A,A _d ,N)	$\sigma_h^2 = \sigma^2 \left[1 + \alpha^2(h-1) + \frac{\beta\phi h}{(1-\phi)^2} \{2\alpha(1-\phi) + \beta\phi\} \right. \\ \left. - \frac{\beta\phi(1-\phi^h)}{(1-\phi)^2(1-\phi^2)} \{2\alpha(1-\phi^2) + \beta\phi(1+2\phi-\phi^h)\} \right]$
(A,N,A)	$\sigma_h^2 = \sigma^2 [1 + \alpha^2(h-1) + \gamma k(2\alpha + \gamma)]$
(A,A,A)	$\sigma_h^2 = \sigma^2 \left[1 + (h-1) \left\{ \alpha^2 + \alpha\beta h + \frac{1}{6}\beta^2 h(2h-1) \right\} \right. \\ \left. + \gamma k \{2\alpha + \gamma + \beta m(k+1)\} \right]$
(A,A _d ,A)	$\sigma_h^2 = \sigma^2 \left[1 + \alpha^2(h-1) + \gamma k(2\alpha + \gamma) \right. \\ \left. + \frac{\beta\phi h}{(1-\phi)^2} \{2\alpha(1-\phi) + \beta\phi\} \right. \\ \left. - \frac{\beta\phi(1-\phi^h)}{(1-\phi)^2(1-\phi^2)} \{2\alpha(1-\phi^2) + \beta\phi(1+2\phi-\phi^h)\} \right. \\ \left. + \frac{2\beta\gamma\phi}{(1-\phi)(1-\phi^m)} \{k(1-\phi^m) - \phi^m(1-\phi^{mk})\} \right]$

Figura 3.1.4: Calcolo della varianza per modello ETS

3.2 Autoregressive Integrated Moving Average

Model

I modelli ARIMA, come il nome suggerisce, si compongono di più sotto-modelli che, assemblati tra di essi, confluiscono in un unico modello finale. I modelli in questione sono moving average models e autoregressive models. La combinazione di essi darà luogo a modelli ARMA, mentre i modelli ARIMA si otterranno quando sarà necessaria una differenziazione dei dati.

3.2.1 Moving Average models

Un modello moving average di ordine q , $MA(q)$, si caratterizza per assumere la seguente forma:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

dove $\varepsilon_t \sim WN(0, \sigma^2)$. y_t è quindi una media ponderata delle stime passate degli errori e le sue proprietà sono le seguenti:

- $E(y_t) = 0$;
- $\text{var}(y_t) = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma^2$;
- $\text{cov}(y_t, y_{t+k}) = \begin{cases} 0, & \text{se } |k| > q, \\ \sigma^2 \sum_{i=0}^{q-|k|} \theta_i \theta_{i+|k|}, & \text{se } |k| \leq q. \end{cases}$

Quest'ultima proprietà ci permette di affermare che un $MA(q)$ è un modello strettamente stazionario.

3.2.2 Autoregressive models

Un modello autoregressivo di ordine p , $AR(p)$, è specificato come:

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t.$$

Il modello in questione è sostanzialmente un modello di regressione dove le variabili indipendenti sono i valori passati della variabile dipendente. Anche in questo caso $\varepsilon_t \sim WN(0, \sigma^2)$.

Considerando le componenti c e ϕ , a seconda dei valori presi da esse, possiamo identificare per un $AR(1)$ varie declinazioni ed andamenti:

- quando $\phi_1 = 0$ e $c = 0$, y_t è una variabile distribuita come una White Noise;
- quando $\phi_1 = 1$ e $c = 0$, y_t è una variabile che segue un andamento di Random Walk;
- quando $\phi_1 = 1$ e $c \neq 0$, y_t è una variabile che segue un andamento di Random Walk con drift;

- quando $\phi_1 = <1$, y_t tende ad oscillare attorno alla media (H., 2010).

Affinché un modello autoregressivo sia stazionario occorre che:

- $-1 < \phi_1 < 1$ per un AR(1);
- $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$, $\phi_1 - \phi_2 < 1$ per un AR(2);

Restrizioni più complesse sono previste per un AR con $p > 2$.

3.2.3 Stazionarietà, differenziazione e random walk model

Nel paragrafo precedente è citato il random walk model. Prima di arrivare al concetto sottostante a questo modello è utile andare a definire cosa si intende per differenziazione delle serie storiche e come ciò si leghi alla stazionarietà di una serie storica.

Iniziando con ordine, una serie storica è definita come stazionaria quando le sue proprietà statistiche non dipendono dal periodo nel quale vengono osservate. Di conseguenza, serie storiche con trend e stagionalità non sono stazionarie perché influenzano i valori a seconda del momento nel quale li osserviamo, mentre serie storiche che seguono una distribuzione white noise lo sono.

Per verificare la presenza di stazionarietà, oltre alla funzione di autocorrelazione, è possibile utilizzare il KPSS test, uno tra i test più utilizzati in questo ambito e fornito dai principali software statistici. In questo test l'ipotesi nulla sostiene che i dati siano stazionari, viceversa per l'ipotesi alternativa. Di conseguenza, valori elevati del test condurranno al rifiuto dell'ipotesi nulla, mentre valori piccoli indicheranno la presenza di non stazionarietà e quindi la necessità di differenziare.

La differenziazione è una metodologia utilizzata per cercare di ottenere dati stazionari, stabilizzando la media e riducendo/eliminando la presenza di trend e stagionalità. Infatti, i dati di una serie storica in valore assoluto potrebbero non essere stazionari, ma le differenze sì.

Una serie storica differenziata è definita come: $y'_t = y_t - y_{t-1}$ dove y'_t ha $T-1$ valori. Questa è la differenziazione di primo ordine. Se una sola differenziazione non fosse sufficiente a rendere la serie storica stazionaria, potrebbe essere necessaria una seconda differenziazione chiamata di secondo ordine, formalmente definita come:

$$\begin{aligned} y''_t &= y'_t - y'_{t-1} \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2} \end{aligned}$$

dove y''_t ha T-2 valori. La differenziazione può essere applicata anche per dati stagionali. La differenza sarà eseguita tra un'osservazione e l'osservazione precedente della stessa stagione/periodo: $y'_t = y_t - y_{t-m}$.

Ora, se la prima differenziazione, ovvero $y_t - y_{t-1}$ segue una distribuzione white noise, allora potrà essere riscritta come $y_t - y_{t-1} = \varepsilon_t$. Riscrivendo il tutto, esplicitando quindi y_t , si otterrà il random walk model: $y_t = y_{t-1} + \varepsilon_t$. Questi modelli hanno dati che evidenziano apparenti lunghi periodi con trend in crescita o decrescita ed improvvisi cambi di direzione.

Se invece il modello non dovesse assumere che le differenze abbiano media zero, identificando questa media differente da zero come c, si ottiene il random walk model con drift: $y_t = c + y_{t-1} + \varepsilon_t$ dove c è la media dei cambiamenti delle osservazioni e, nel caso in cui sia positiva, il valore di y_t crescerà nel tempo, viceversa se i valori di c sono negativi.

3.2.4 Invertibilità

Un'importante proprietà tra i modelli moving average e i modelli autoregressivi è l'invertibilità. Ogni modello AR(p) stazionario può essere riscritto in un modello MA(∞). Considerando un modello AR(1):

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \varepsilon_t \\ &= \phi_1 (\phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \phi_1^2 y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &= \phi_1^3 y_{t-3} + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &\quad \dots \end{aligned}$$

dato che ϕ_1 sarà compreso tra -1 ed 1 essendo AR stazionario, il valore di ϕ_1 diventerà sempre più piccolo all'aumentare dell'esponente ottenendo quindi:

$$y_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} \dots$$

ovvero, un modello MA(∞).

L'invertibilità si ottiene quando anche un modello MA(q) può essere riscritto come un modello AR(∞). In un modello MA(1) $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$ il termine di errore ε_t può essere riscritto come una funzione lineare delle correnti e passate osservazioni di y_t :

$$\varepsilon_t = \sum_{j=0}^{\infty} (-\theta_1)^j y_{t-j}.$$

L'invertibilità è presente quando $|\theta_1| < 1$ così da denotare il maggior peso delle osservazioni più recenti rispetto alle passate. Questa restrizione è valida per modelli MA(1), per modelli MA(2) le restrizioni saranno: $-1 < \theta_2 < 1$, $\theta_1 + \theta_2 < -1$, $\theta_1 - \theta_2 < 1$. Restrizioni più complesse sono previste per modelli MA con $q > 2$.

3.2.5 ARMA model

La necessità della creazione di un modello che mettesse insieme moving average e autoregressioni nasce dalla necessità di avere un modello che non necessiti di lunghe serie di dati per compiere il proprio obiettivo e questo sarebbe stato necessario utilizzando solamente AR piuttosto che MA. Y_t è un processo ARMA(p,q) se:

- y_t è stazionario;
- per tutti i tempi t , $\phi(B)Y_t = \theta(B)\varepsilon_t$ dove B , in gergo chiamata backshift notation, è uguale rispettivamente a y_{t-1} e ε_{t-1} .

Per i modelli ARMA assumeremo che essi siano invertibili e casuali (dove per casualità si intende che nel processo autoregressivo gli esponenti di ϕ siano tutti maggiori di 1).

3.2.6 ARIMA model

Dal modello ARMA, si passa al modello ARIMA quando i dati vengono differenziati. Un modello ARIMA(p,d,q), dove la lettera d indica il grado di differenziazione (solitamente uguale o inferiore a 3), è rappresentato come:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_q.$$

A seconda dell'ordine di processi autoregressivi, del grado di differenziazione e dell'ordine della moving average, si possono ottenere vari modelli ARIMA, tra cui alcuni casi speciali:

- ARIMA(0,0,0) senza costante è un processo di white noise;
- ARIMA(0,1,0) senza costante è un processo random walk;
- ARIMA(0,1,0) con costante è un processo random walk con drift;
- ARIMA(p,0,0) è un modello autoregressivo;
- ARIMA(0,0,q) è un modello moving average.

È importante notare che, a seconda del valore della costante all'interno del modello e del grado di differenziazione che viene scelto, si avranno effetti diversi sulle stime future. In particolare:

- se $c = 0$ e $d = 0$, le stime future di lungo periodo andranno a 0;
- se $c = 0$ e $d = 1$, le stime future di lungo periodo tenderanno ad una costante diversa da 0;
- se $c = 0$ e $d = 2$, le stime future di lungo periodo tenderanno ad una linea retta;
- se $c \neq 0$ e $d = 0$, le stime future di lungo periodo convergeranno alla media dei dati;
- se $c = 0$ e $d = 1$, le stime future di lungo periodo seguiranno una linea retta;

- se $c = 0$ e $d = 2$, le stime future di lungo periodo seguiranno un trend quadratico.

Il valore di d ha inoltre effetti sugli intervalli di stima futuri, infatti più alti saranno i valori di d , maggiori saranno gli incrementi in dimensione degli intervalli, viceversa per valori piccoli o nulli di d . In quest'ultimo caso gli intervalli di stima futuri saranno sostanzialmente identici. Il valore che determina l'ordine del processo autoregressivo ha invece effetti se nei dati sono presenti cicli. Se si vogliono ottenere stime future con un andamento ciclico, è necessario avere $p \geq 2$.

3.2.7 Seasonal ARIMA

Un modello ARIMA che presenti dati con stagionalità e che sia in grado di stimarli anche in futuro è denominato come SARIMA(p,d,q)x(P,D,Q) $_m$ dove la seconda parte del modello indica la parte che cattura la stagionalità ed m indica il periodo dell'anno.

Formalmente il modello può essere scritto come:

$$\phi(B)\Phi_P(B^m)(1-B)^d(1-B^m)^D y_t = \theta(B)\Theta_Q(B^m)\varepsilon_t$$

dove

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p,$$

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q,$$

$$\Phi_P(B^m) = 1 - \Phi_1 B^m - \dots - \Phi_P B^{m^P},$$

$$\Theta_Q(B^m) = 1 - \Theta_1 B^m - \dots - \Theta_Q B^{m^Q}.$$

3.2.8 Funzione di autocorrelazione e funzione di autocorrelazione parziale

Ottenere grafici delle autocorrelazioni dei dati risulta utile quando si scelgono i parametri dei modelli ARIMA stagionali e non. Come descritto precedentemente, la funzione di autocorrelazione evidenzia la relazione tra i valori di y_t e y_{t-k} per differenti valori di k . La necessità della funzione di autocorrelazione parziale nasce dal fatto che se y_t e y_{t-1} sono correlati, allora anche y_{t-1} e y_{t-2} devono esserlo, ma y_t e y_{t-2} non è detto che lo siano e potrebbero esserlo solamente perché y_{t-1} li collega. La funzione di autocorrelazione parziale permette di rimuovere gli effetti dei valori ritardati per $k = 1, 2, 3, \dots, k-1$. Dunque, la prima autocorrelazione parziale è uguale alla prima autocorrelazione dato che non ci sono valori ritardati intermedi da rimuovere.

L'aiuto dei grafici che mostrano autocorrelazione e autocorrelazione parziale avviene solamente nel caso in cui entrambi i valori p e q non siano positivi.

Quando i dati mostrano:

- una ACF esponenzialmente decadente o con un andamento sinusoidale;
- un picco al lag p nella PACF, ma nessun altro dopo tale lag,

allora un modello ARIMA(p,d,0) è adatto a rappresentare correttamente i dati.

Mentre se i dati mostrano:

- una PACF esponenzialmente decadente o con un andamento sinusoidale;
- un picco al lag q nella ACF, ma nessun altro dopo tale lag,

allora un modello ARIMA(0,d,q) è adatto a rappresentare correttamente i dati.

3.2.9 Selezione del modello

Una volta identificati gli ordini p, d e q, occorre scegliere i valori dei parametri del modello, i quali sono automaticamente scelti da vari software statistici utilizzando la stima della massima probabilità, tecnica che consente di trovare i parametri che più permetteranno al modello di creare dati che si avvicinino ai dati osservati.

Una volta che il modello è definito, la selezione dei vari modelli avviene anche in questo caso attraverso tre indicatori:

- Akaike's Information Criterion, calcolato come:

$$AIC = -2 \log(L) + 2(p + q + k + 1)$$

dove L è la probabilità associata ai dati e k = 1 se c ≠ 0 oppure k = 0 se c = 0.

- Akaike's Information Criterion corretto, calcolato come:

$$AIC_c = AIC + \frac{2(p + q + k + 1)(p + q + k + 2)}{T - p - q - k - 2}$$

- Bayesian Information Criterion, definito dalla formula:

$$BIC = AIC + [\log(T) - 2](p + q + k + 1)$$

In tutti i casi, il modello migliore sarà quello che minimizza i vari criteri.

3.2.10 Stime future

Per quanto riguarda le stime puntuali future, il processo avviene semplicemente esplicitando y_t , riscrivendo l'equazione rimpiazzando a t, T+h, e infine sostituendo nella parte destra dell'equazione le osservazioni future con le loro stime, gli errori futuri con 0, e li errori passati con i residui generati.

Gli intervalli di stima sono calcolati grazie all'utilizzo di software statistici.

Capitolo 4 - Applicazione dei modelli

In questo capitolo andremo ad utilizzare i dati del brand H dell'area Q dell'azienda Safilo per applicare la teoria dei metodi di previsione semplici visti nel capitolo 2 e dei modelli exponential smoothing e ARIMA. I metodi semplici serviranno come benchmark per i modelli più complessi per confrontare la reale necessità di utilizzare tecniche più complesse nella previsione delle vendite.

Nella figura 4.1 sono rappresentate le vendite trimestrali del Brand H.

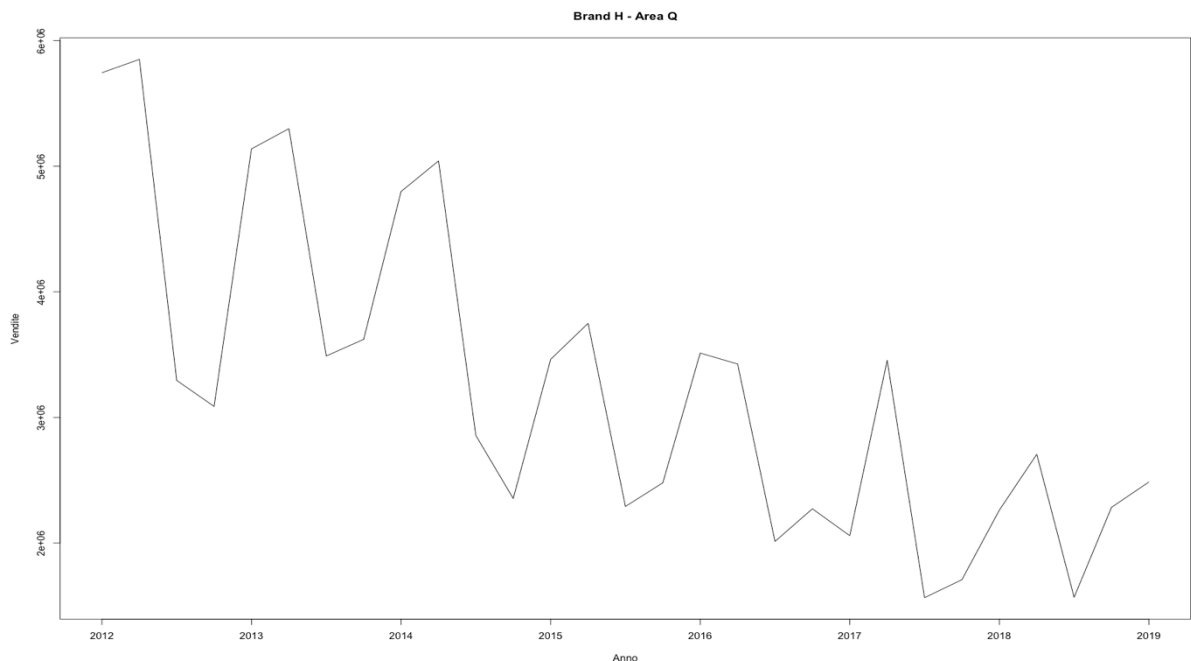


Figura 4.1: Vendite trimestrali Brand H – Area Q

Guardando attentamente il grafico, è chiara la presenza di un trend decrescente dei dati oltre ad una spiccata componente stagionale. Il grafico mette chiaramente in mostra la presenza di autocorrelazione dei dati senza che sia necessario ricorrere alle funzioni di autocorrelazione/autocorrelazione parziale stimate.

Procediamo quindi ad effettuare le previsioni utilizzando i metodi base mean, naïve, seasonal naïve e drift. Per effettuare ciò ci serviremo del software R utilizzando i pacchetti “forecast” ed “fpp3” per effettuare le previsioni future e del pacchetto “tidyverse” per plottare automaticamente un grafico che si adatti ai modelli che vogliamo esporre.

Prima di effettuare le previsioni, è necessario andare a dividere i nostri dati di vendita in training data e test data per verificare successivamente l'accuratezza delle previsioni dei modelli. I training set vanno dal primo trimestre del 2012 al quarto trimestre del 2016, mentre

i test data si compongono delle vendite dal primo trimestre del 2017 al quarto trimestre del 2018.

Nella figura 4.2 sono dunque rappresentate le vendite consuntive del Brand H con la linea nera, mentre le altre linee raffigurano le vendite stimate a seconda del metodo come specificato in didascalia.

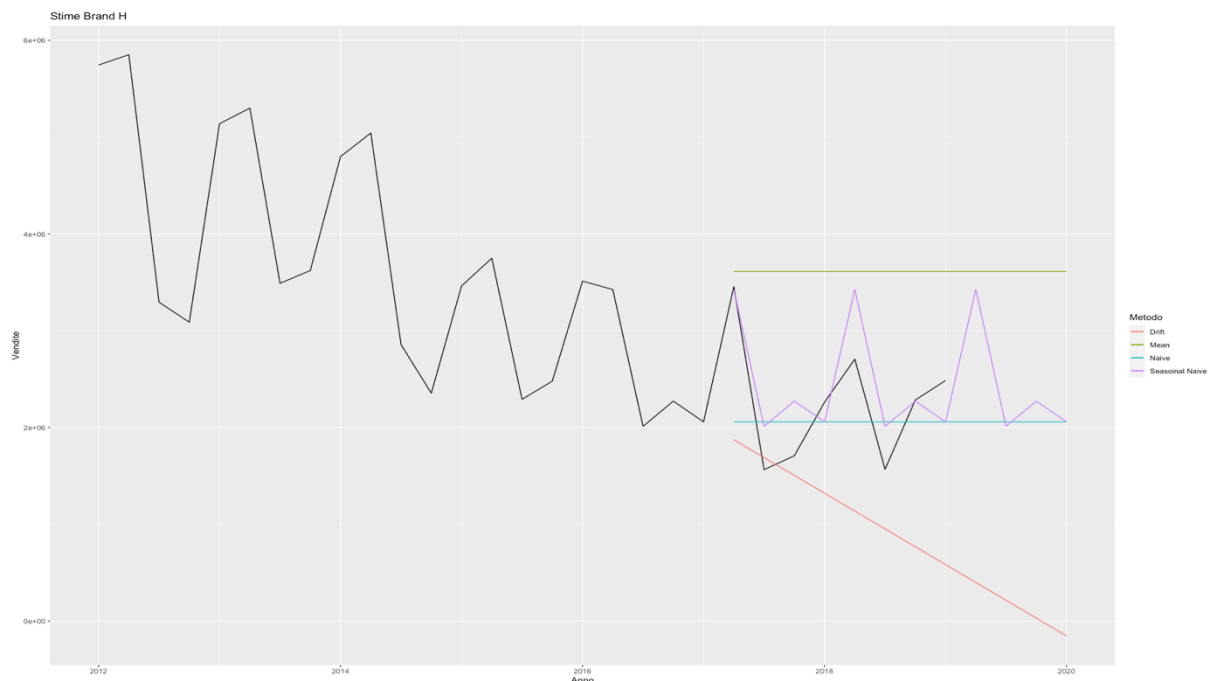


Figura 4.2: Stime di vendita Brand H – Area Q con metodi mean, naïve, seasonal naïve e drift

Come prevedibile, già dal grafico si intuisce come il seasonal naïve catturi meglio l'andamento dei dati. Andando nel dettaglio, attraverso la funzione `accuracy()`, è possibile calcolare l'accuratezza delle previsioni basandoci sulla differenza tra dati stimati e test data. Nella figura 4.3 sono illustrati i risultati di tale calcolo.

```
> accuracy(fc_mean_brand_h_q, brand_h_q_test)
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1 Theil's U
Training set -2.209856e-11 1207316  985898.3 -11.69081 30.16904 1.673357 0.4413732    NA
Test set     -1.350212e+06 1403195 1350212.3 -65.35774 65.35774 2.291705 -0.4374487 1.978112
> accuracy(fc_naive_brand_h_q, brand_h_q_test)
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1 Theil's U
Training set -184293.8 1167003.9 858939.0 -10.744291 27.33032 1.4578705 -0.04444755    NA
Test set      202745.3 432420.4 398844.2  5.745489 18.24897 0.6769552 -0.43744868 0.6286221
> accuracy(fc_snaive_brand_h_q, brand_h_q_test)
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1 Theil's U
Training set -482872.5 740338.5 589173.8 -16.727006 19.87753 1.0000000 0.47065269    NA
Test set     -104188.1 433090.1 360955.6 -5.640327 16.30930 0.6126471 0.04784115 0.6144456
> accuracy(fc_drift_brand_h_q, brand_h_q_test)
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1 Theil's U
Training set -6.986962e-11 1152360  887036.7 -4.920273 27.17103 1.50556 -0.04444755    NA
Test set      1.308508e+06 1388505 1308508.0 56.331245 56.33124 2.22092 -0.27581172 2.237075
```

Figura 4.3: Accuratezza previsiva metodi mean, naïve, seasonal naïve e drift

I valori da osservare sono quelli in corrispondenza della riga dei test set. I valori del RMSE ci confermano che naïve e seasonal naïve sono i metodi che minimizzano la differenza tra test data e previsioni. Considerando altri indicatori come MAPE e MASE, il seasonal naïve ottenendo i valori più bassi sembrerebbe essere quello migliore.

Nei capitoli 4.1 e 4.2 si effettueranno le previsioni con i modelli ETS e ARIMA per tentare di ottenere stime migliori rispetto a quelle ottenute con i metodi precedenti.

4.1 Applicazione dell'exponential smoothing model

Attraverso i seguenti codici in R rappresentati in figura 4.1.1, il software sceglierà automaticamente il modello che più si adatta ai dati.

```
> ets_brand_h_q = ets(brand_h_q_training,
+                      alpha = NULL,
+                      beta = NULL,
+                      gamma = NULL,
+                      phi = NULL)
> ets_brand_h_q
ETS(M,N,M)

Call:
ets(y = brand_h_q_training, alpha = NULL, beta = NULL, gamma = NULL,

Call:
  phi = NULL)

Smoothing parameters:
  alpha = 0.9271
  gamma = 1e-04

Initial states:
  l = 4343276.9729
  s = 0.8074 0.7694 1.2463 1.1769

sigma: 0.1368

      AIC      AICc      BIC
621.0694 629.6848 628.3811
```

Figura 4.1.1: Parametri modello ETS

Il modello scelto è dunque un ETS(M,N,M) colui il quale minimizza i valori di AIC, AICc e BIC. Andiamo ora a verificare se il modello ottenuto ha residui autocorrelati o meno con le funzioni di autocorrelazione (Figura 4.1.2) e di autocorrelazione parziale (Figura 4.1.3).

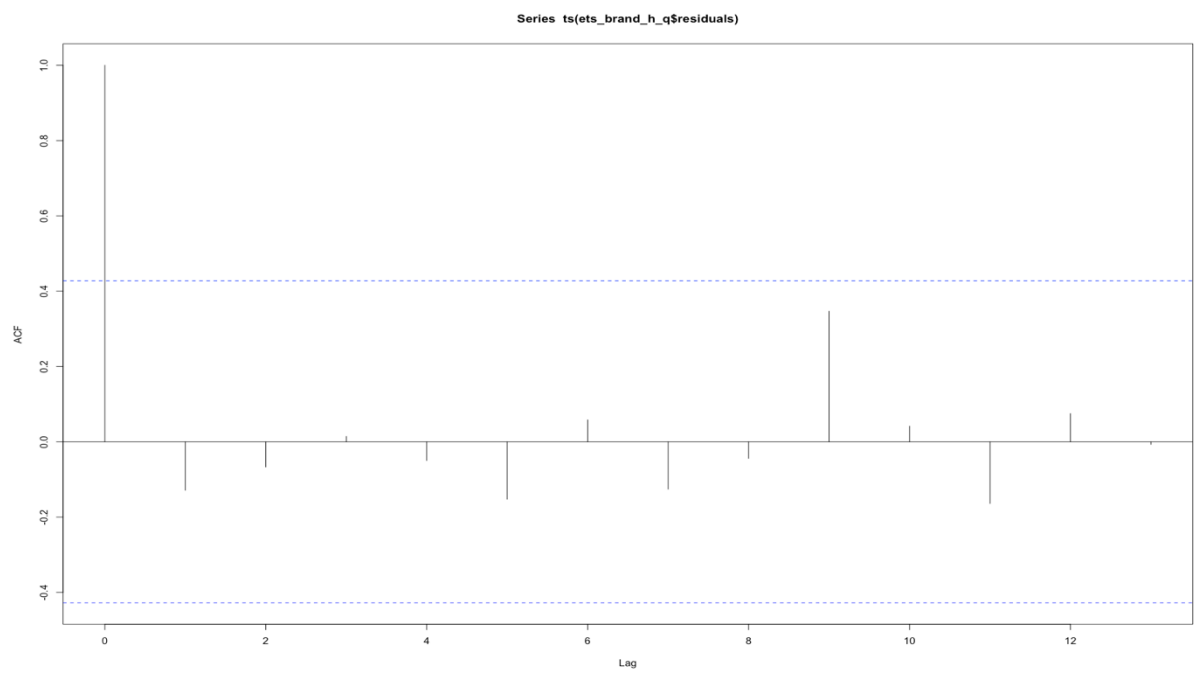


Figura 4.1.2: Funzione di autocorrelazione modello ETS

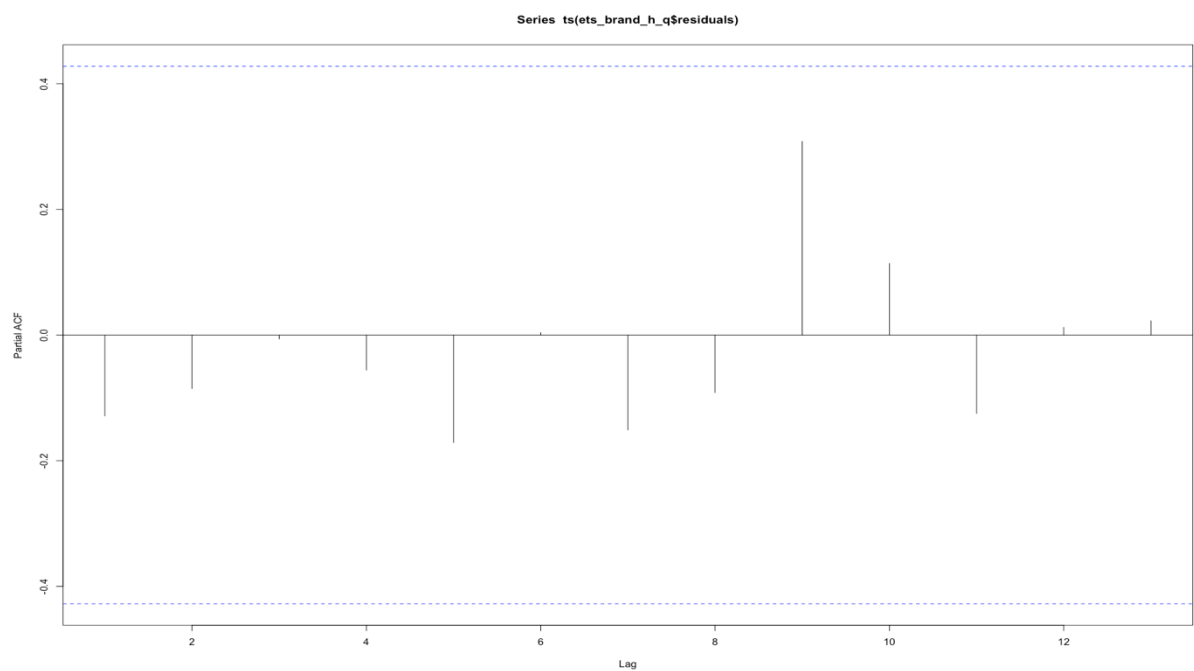


Figura 4.1.3: Funzione di autocorrelazione parziale modello ETS

Come evidenziano le immagini, sia la ACF, sia la PACF non evidenziano residui autocorrelati.

A questo punto non resta che stimare gli andamenti futuri. Attraverso i comandi rappresentati in figura 4.1.4, abbiamo stimato le previsioni di vendita per 3 anni (12 trimestri) ad un livello di confidenza sia dell'80, che del 95%. La prima colonna indica le stime puntuali, mentre le successive i valori inferiori e superiori degli intervalli per trimestre.

```
> fc_ets_brand_h_q = forecast(ets_brand_h_q, level = c(80, 95), h = 12)
> fc_ets_brand_h_q
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q2	2275592	1876637.5	2674547	1665443.5	2885741
2017 Q3	1404840	1067525.7	1742154	888962.4	1920717
2017 Q4	1474161	1044371.4	1903950	816854.8	2131466
2018 Q1	2148727	1427097.7	2870356	1045090.3	3252364
2018 Q2	2275597	1421018.3	3130175	968631.8	3582562
2018 Q3	1404842	826087.9	1983597	519713.8	2289971
2018 Q4	1474163	816760.3	2131566	468752.1	2479575
2019 Q1	2148731	1121620.8	3175841	577901.4	3719561
2019 Q2	2275601	1118406.9	3432795	505825.3	4045377
2019 Q3	1404845	649388.5	2160302	249473.9	2560216
2019 Q4	1474166	639927.7	2308405	198308.6	2750024
2020 Q1	2148735	874222.7	3423247	199536.6	4097934

Figura 4.1.4: Stime di vendita puntuali ed intervalli di confidenza modello ETS

Graficamente:

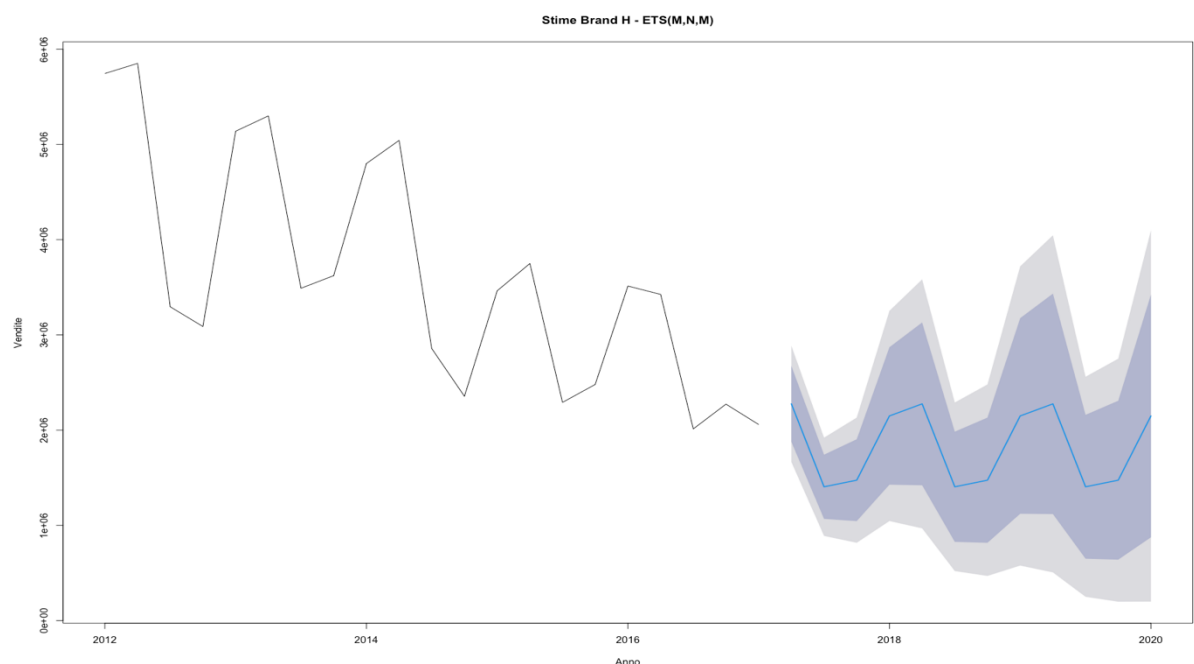


Figura 4.1.5: Grafico stime di vendita modello ETS

Ottenute le stime, effettuiamo nuovamente una verifica di come si distribuiscono i residui.

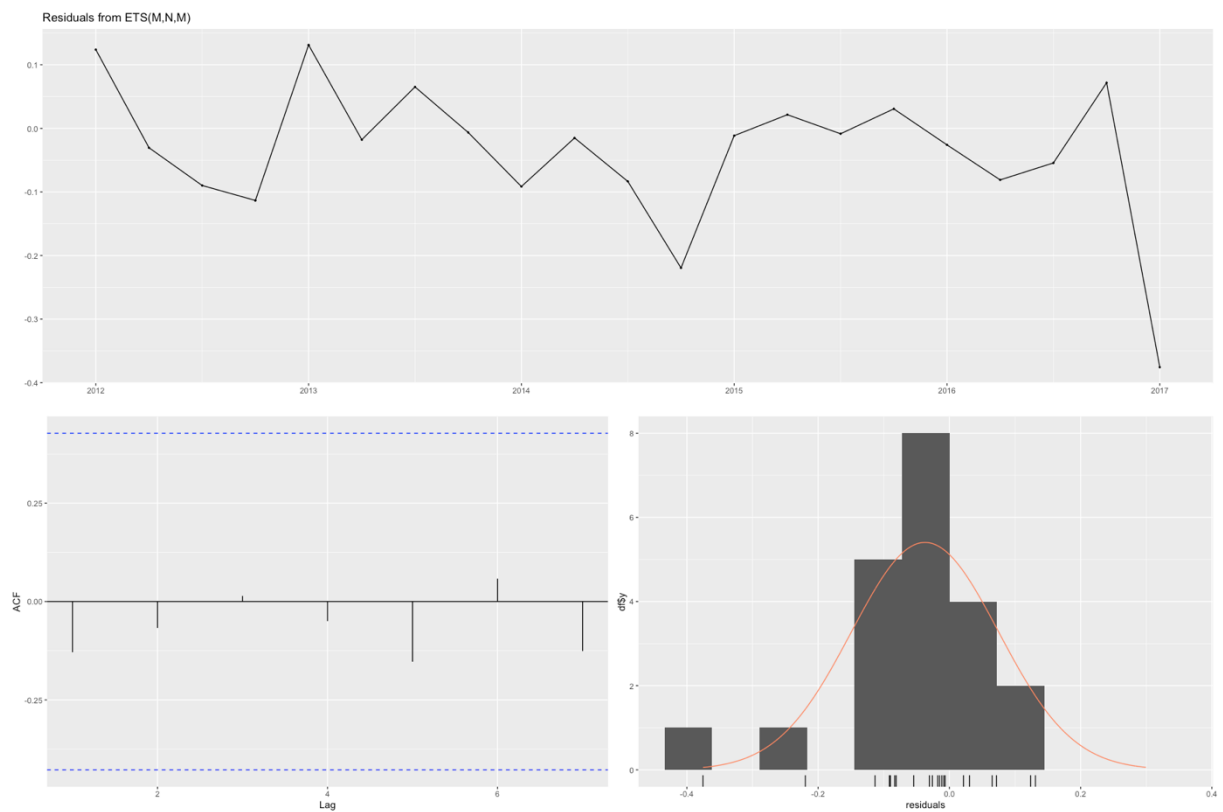


Figura 4.1.6: Analisi dei residui modello ETS

```
> checkresiduals(fc_ets_brand_h_q)

Ljung-Box test

data: Residuals from ETS(M,N,M)
Q* = 6.8634, df = 3, p-value = 0.07638

Model df: 6. Total lags used: 9
```

Figura 4.1.7: Test di Ljung-Box modello ETS

Le immagini in figura 4.1.6 rappresentano chiaramente la distribuzione normale di essi e la mancanza di autocorrelazione. Tutto ciò è confermato anche dal test di Ljung-Box (figura 4.1.7) il quale ottiene un p-value che ci permette di non rifiutare l'ipotesi di stazionarietà.

4.2 Applicazione del modello ARIMA

Anche per il modello ARIMA, il software andrà a scegliere i parametri che meglio si adattano ai dati. La scelta del modello in base a quale criterio, potrà essere specificata nel codice. In questo caso (figura 4.2.1), AIC, AICc e BIC hanno portato alla scelta dello stesso modello, ovvero:

```
> arima_brand_h_q
Series: brand_h_q_training
ARIMA(0,0,1)(1,1,0)[4] with drift

Coefficients:
          ma1          sar1          drift
          0.8489      -0.5759     -137196.82
s.e.      0.2622       0.2357      27532.27

sigma^2 = 1.57e+11:  log likelihood = -243.3
AIC=494.6   AICc=497.94   BIC=497.93
```

Figura 4.2.1: Parametri modello ARIMA

Ovviamente il software ha catturato la stagionalità presente nei dati, stiamo quindi parlando di un modello seasonal ARIMA.

Effettuiamo le verifiche sull'autocorrelazione dei residui con rispettivamente ACF (figura 4.2.2) e PACF (figura 4.2.3):

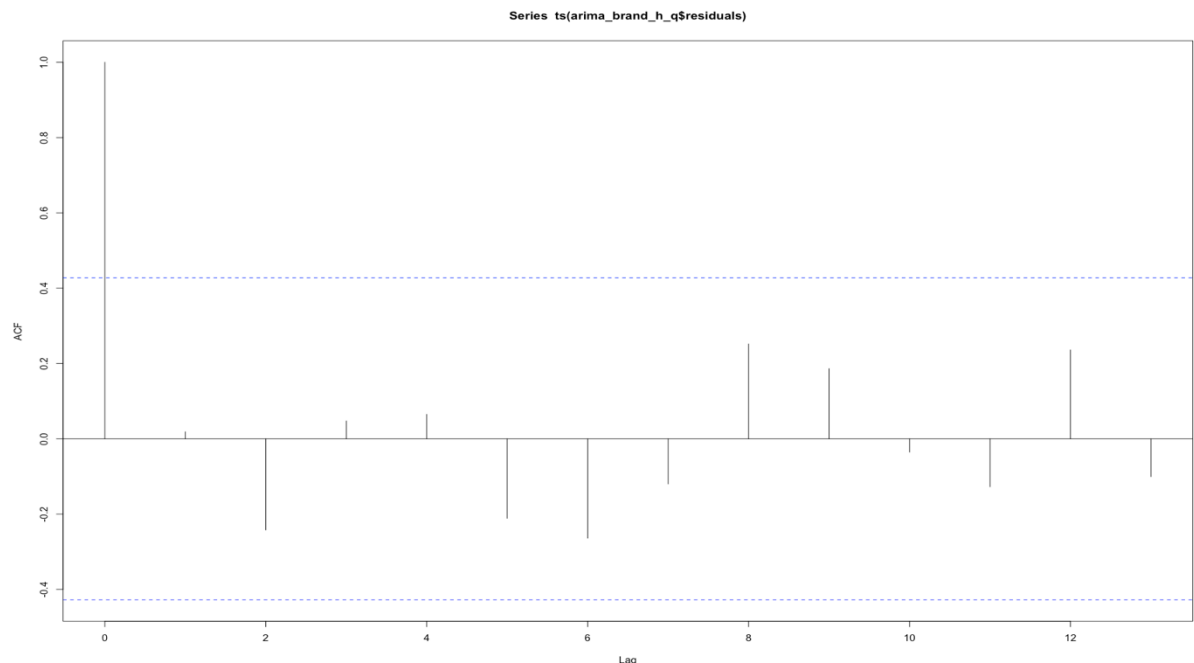


Figura 4.2.2: Funzione di autocorrelazione modello ARIMA

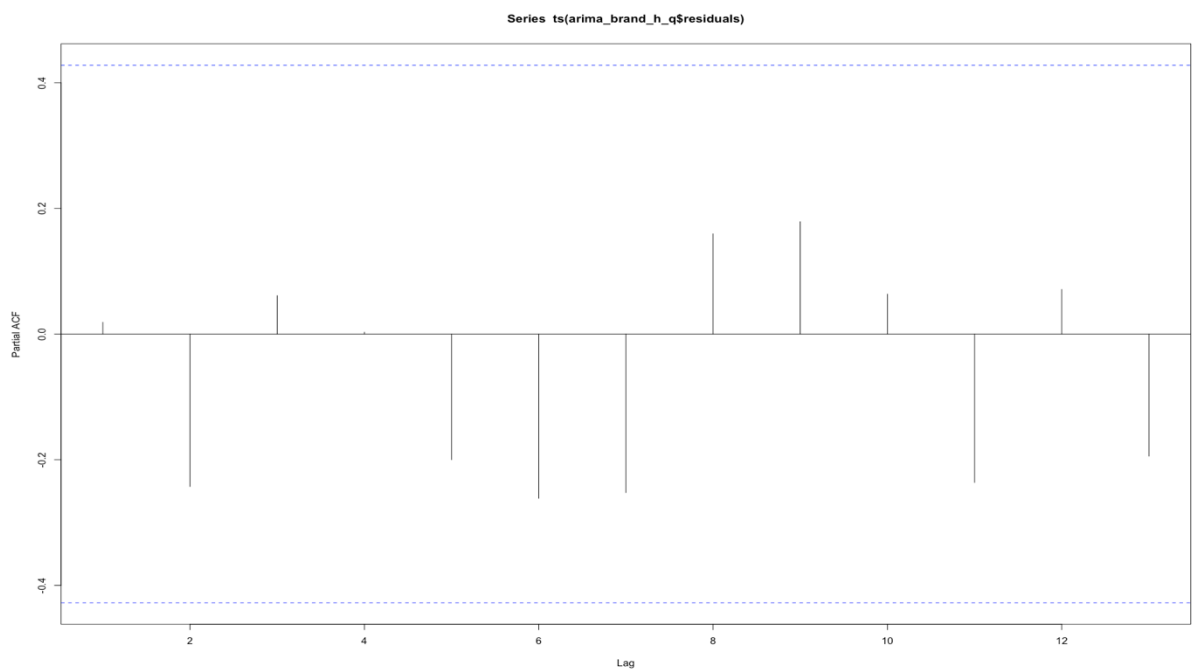


Figura 4.2.3: Funzione di autocorrelazione parziale modello ARIMA

Nessuno dei due grafici evidenzia autocorrelazione.

Passiamo quindi alla previsione delle vendite per i quattro anni successivi (figura 4.2.4).

```
> fc_arima_brand_h_q = forecast(arima_brand_h_q, level = c(80, 95), h=12)
> fc_arima_brand_h_q
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q2	1992467.2	1484095.93	2500839	1214980.42	2769954
2017 Q3	1308128.9	641962.91	1974295	289315.92	2326942
2017 Q4	1526932.0	860766.00	2193098	508119.02	2545745
2018 Q1	2031306.3	1365140.26	2697472	1012493.28	3050119
2018 Q2	1952802.3	1252622.59	2652982	881969.82	3023635
2018 Q3	848949.0	125360.85	1572537	-257683.60	1955582
2018 Q4	1091613.7	368025.52	1815202	-15018.93	2198246
2019 Q1	1182170.7	458582.49	1905759	75538.04	2288803
2019 Q2	1110790.6	291524.26	1930057	-142169.16	2363750
2019 Q3	248553.2	-632955.84	1130062	-1099598.54	1596705
2019 Q4	477475.0	-404034.02	1358984	-870676.72	1825627
2020 Q1	806366.3	-75142.74	1687875	-541785.44	2154518

Figura 4.2.4: Stime di vendita puntuali e intervalli di confidenza modello ARIMA

Anche in questo caso, i livelli di confidenza scelti sono dell'80 e 95%. I valori degli intervalli di confidenza sono raffigurati nell'immagine. Graficamente:

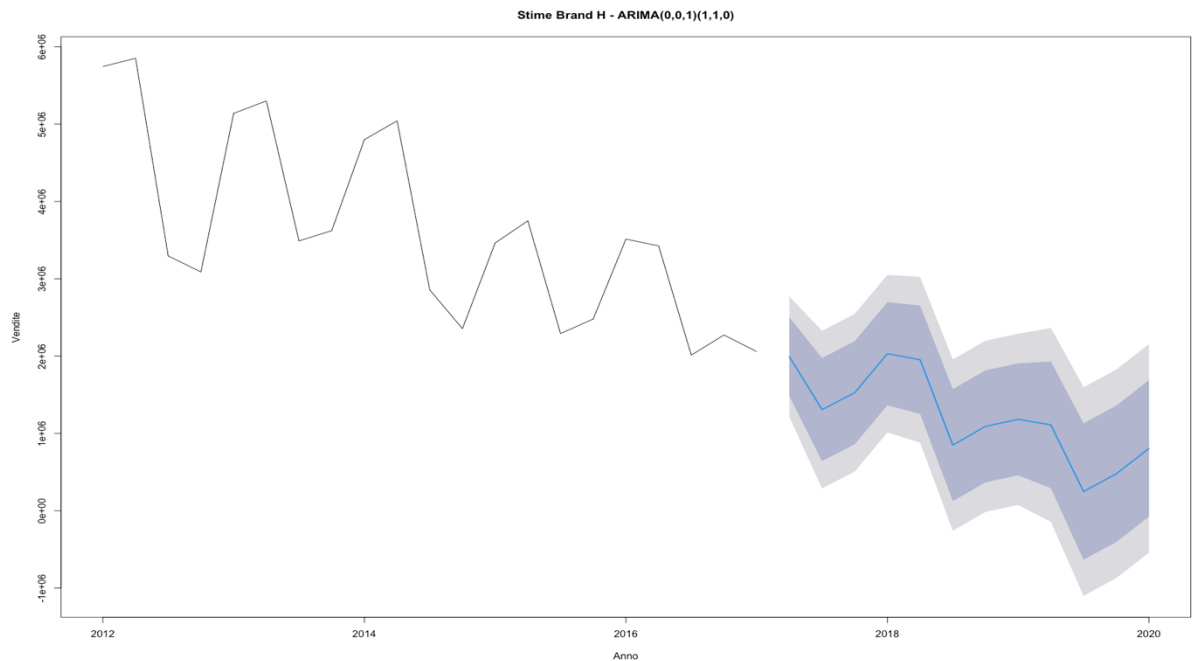


Figura 4.2.5: Grafico stime di vendita modello ARIMA

Arrivati a questo punto, un check dei residui è necessario per verificare come essi si distribuiscono una volta effettuate le stime.

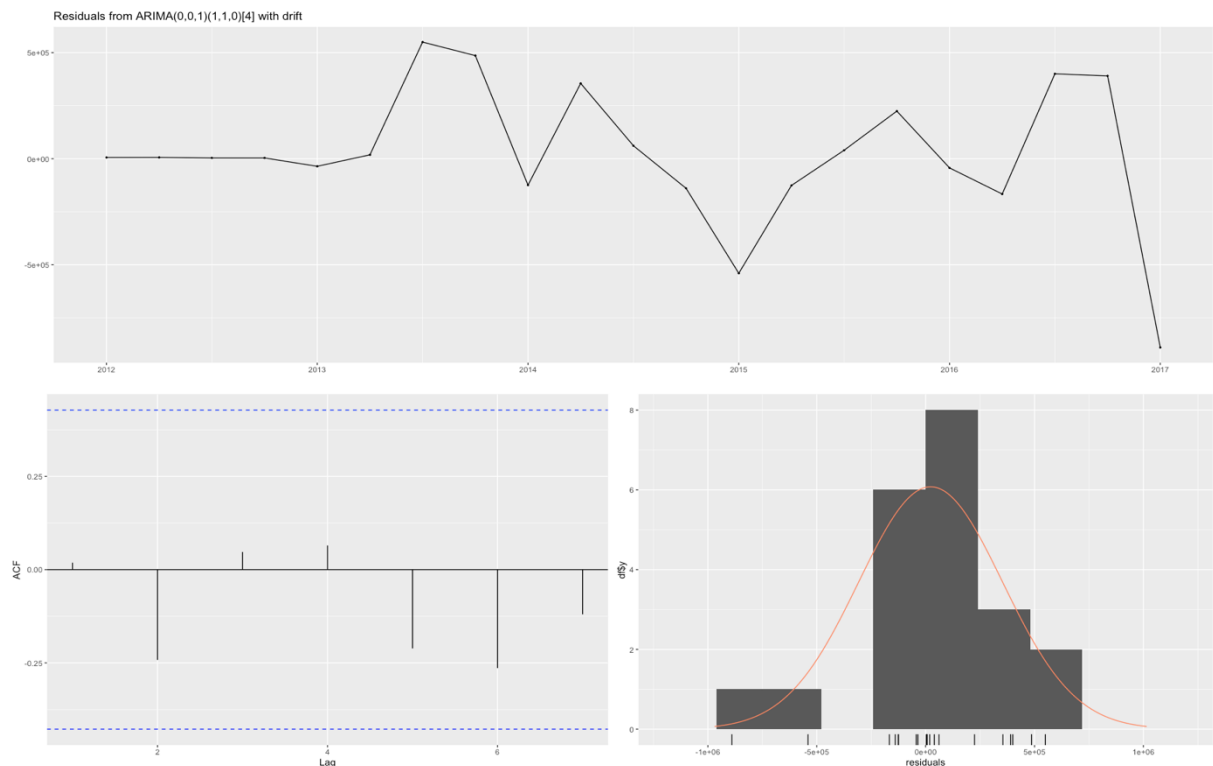


Figura 4.2.6: Analisi dei residui modello ARIMA

```
> checkresiduals(fc_arima_brand_h_q)

Ljung-Box test

data: Residuals from ARIMA(0,0,1)(1,1,0)[4] with drift
Q* = 5.2715, df = 3, p-value = 0.153
```

Figura 4.2.7: Test di Ljung-Box modello ARIMA

Come il residual check (figura 4.2.6) e il test di Ljung Box (figura 4.2.7) evidenziano, possiamo affermare che i residui non sono tra di essi correlati.

4.3 Confronto Seasonal naïve, ETS e ARIMA

A questo punto andiamo a confrontare l'accuratezza delle previsioni del metodo seasonal naïve (il migliore tra i semplici) con quella dell'exponential smoothing e dell'ARIMA.

In figura 4.3.1 sono rappresentati i dati ottenuti.

```
> accuracy(fc_snaive_brand_h_q, brand_h_q_test)
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1 Theil's U
Training set -482872.5 740338.5 589173.8 -16.727006 19.87753 1.0000000 0.47065269      NA
Test set     -104188.1 433090.1 360955.6  -5.640327 16.30930 0.6126471 0.04784115 0.6144456

> accuracy(fc_ets_brand_h_q, brand_h_q_test)
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1 Theil's U
Training set -122302.6 410294.9 288686.4  -5.522505  9.380161 0.4899852 -0.1209664      NA
Test set      370935.2 446119.8 370935.2 16.073996 16.073996 0.6295853 -0.4355988 0.872377

> accuracy(fc_arima_brand_h_q, brand_h_q_test)
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1 Theil's U
Training set  22832.8 323561.9 219548.4  0.4531425  7.833532 0.3726377 0.01865768      NA
Test set      839978.9 923140.4 839978.9 37.7121075 37.712108 1.4256895 0.25102087 1.585939
```

Figura 4.3.1: Accuratezza previsiva seasonal naïve, ETS e ARIMA

Attraverso questi risultati, il modello ARIMA non risulta essere appropriato in quanto ottiene un RMSE, un MAPE superiori rispetto agli altri due modelli. Concentrandoci sul confronto tra seasonal naïve e exponential smoothing, essi ottengono valori poco distanti sia in termini di RMSE, che in termini di MAPE.

Queste deduzioni sono chiare anche dal grafico esposto in figura 4.3.2.

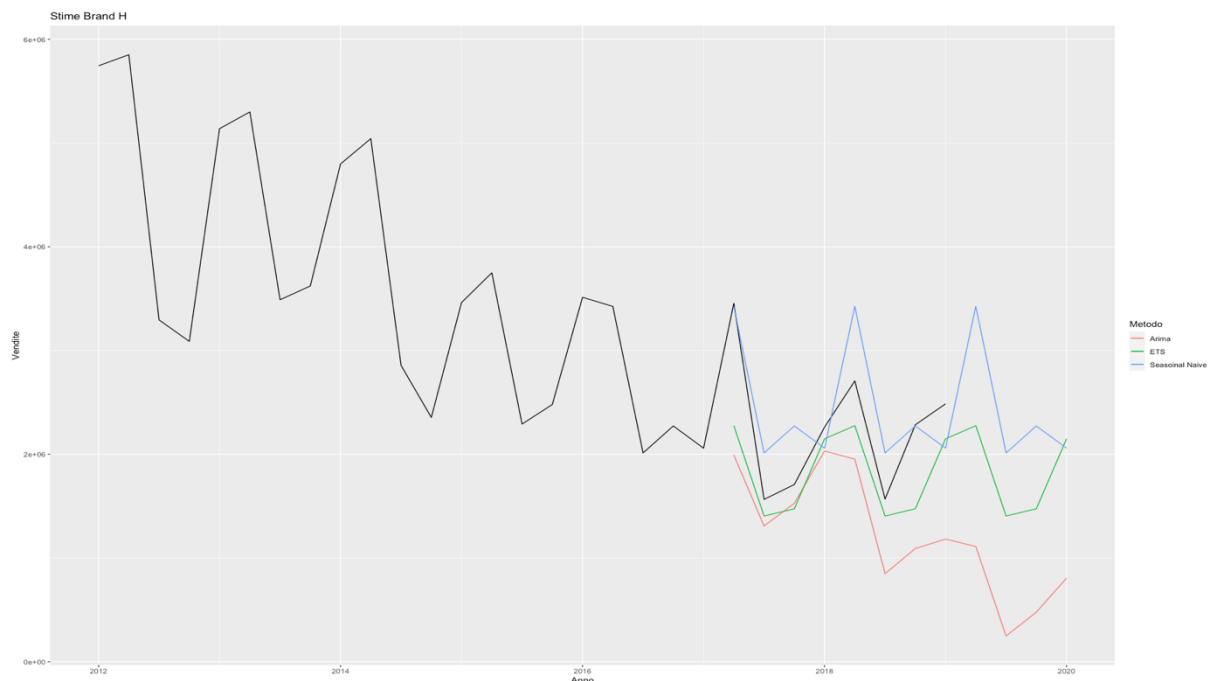


Figura 4.3.2: Grafico stime di vendita seasonal naïve, ETS e ARIMA

Conclusioni

L'analisi dei dati, ed ancor di più la loro previsione futura, oggi sono sempre più oggetto di interesse ed applicazione nel mondo aziendale. A tutto ciò si affianca il mondo dell'informatica e più precisamente della programmazione, la quale permette di poter mettere in pratica svariate tecniche di analisi, da quelle classiche, sino a quelle più moderne.

Ecco che in questo documento si è passati dalla spiegazione delle tecniche più semplici di previsione futura, alcune delle quali ovviamente offrono una visione delle stime future piuttosto irrealistiche e semplici, ai modelli più complessi che invece permettono di adattarsi alla specifica fattispecie presa in considerazione.

I modelli liscamente esponenziali e ARIMA, infatti, permettono una previsione sulla base dei dati passati e delle loro caratteristiche. Come visto nell'ultimo capitolo, però, non sempre modelli più complessi portano a stime più precise: il capitolo 4.3 ha messo in luce come un modello come l'ARIMA non abbia fornito risultati migliori rispetto ad una tecnica come il seasonal naïve, ed anche il modello liscamente esponenziale non ha ottenuto stime significativamente migliori rispetto alla tecnica citata.

Chiaramente il settore preso in considerazione si presta a forte incertezza derivante da una moltitudine di variabili, il che rende la previsione delle vendite future di occhiali, ovvero lo scopo di questa tesi, molto ambizioso. Queste tecniche sono però alla base dei modelli attualmente utilizzati in tali contesti.

Bibliografia

Dagum E. B., Bianconcini S. Seasonal adjustment methods and real time trend-cycle estimation [Libro]. - [s.l.] : Springer, 2016.

Evans M. K. Practical Business Forecasting [Libro]. - Malden : Blackwell Publishers, 2002.

Chan N. H. Time Series: Applications to Finance with R and S-Plus®, Second Edition [Libro]. - Hoboken : John Wiley & Sons, 2010.

Holt C. Forecasting seasonals and trends by exponentially weighted averages [Rivista]. - [s.l.] : International Journal of Forecasting, 2004.

Hyndman R. J., Koehler A. B. Another look at measures of forecast accuracy [Rivista] // International Journal of Forecasting. - [s.l.] : Elsevier, 2006. - 4 : Vol. 22. - p. 679-688.

Hyndman R. J., Koehler A. B., Ord J. K., Snyder, R. D. Forecasting with Exponential Smoothing: The State Space Approach [Libro]. - [s.l.] : Springer-Verlag, 2008.

Hyndman R.J., Athanasopoulos G. Forecasting: principles and practice, 3rd edition [Libro]. - Melbourne : OTexts, 2021.

Guidolin M., Pedio M. Essential of Time Series for Financial Applications [Libro]. - [s.l.] : Elsevier, 2018. - p. Cap. 2.

Montgomery D. C., Jennings C. L., Kulahci M. Introduction to Time Series Analysis and Forecasting [Libro]. - Hoboken, New Jersey : John Wiley & Sons, 2015.

Winters P. R. Forecasting Sales by Exponentially Weighted Moving Averages [Rivista]. - [s.l.] : Management Science, 1960. - 3 : Vol. 6.