

Previsioni di Vendita mediante Serie Storiche

Edoardo Ruffoli

27 Dicembre 2021

Indice

1	Introduzione	1
2	Dataset	1
3	Analisi Preliminare	1
4	Decomposizione	3
4.1	Decomposizione Stagionalità Stazionaria	3
4.2	Decomposizione STL	3
4.3	Analisi dei Residui	3
5	Previsione	5
5.1	Holt Winters	5
5.2	Metodi Autoregressivi	6
5.3	Analisi dei Residui	6
5.4	Confronto	7
6	Conclusioni	8

1 Introduzione

Si ipotizza che la seguente analisi sia stata commissionata da un'azienda automobilistica con lo scopo di riuscire a prevedere l'andamento delle vendite di automobili nei prossimi anni.

Una previsione affidabile del trend è fondamentale per valutare quali siano i migliori momenti in cui aumentare la produzione; inoltre la conoscenza della stagionalità delle vendite permetterebbe all'azienda di pianificare le campagne pubblicitarie e le sponsorizzazioni.

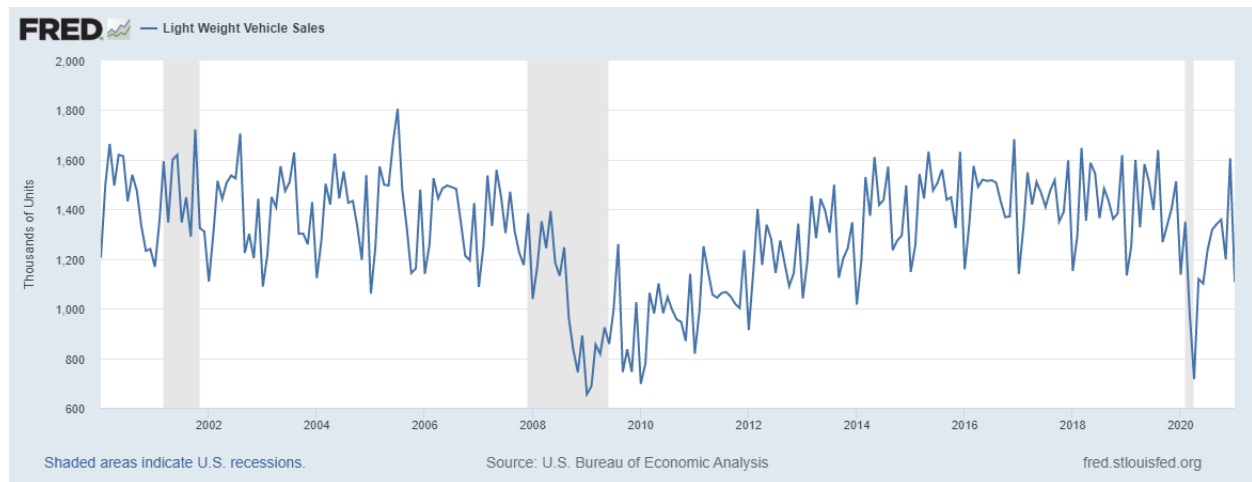
2 Dataset

L'analisi è stata svolta utilizzando il seguente dataset: <https://fred.stlouisfed.org/series/LTOTALNSA>. Esso contiene una serie storica con osservazioni mensili riguardanti il numero di vendite negli Stati Uniti di *light weight veicles*, ovvero veicoli con un peso inferiore alle 10'000 libbre. Per l'analisi sono state considerate solo le osservazioni dal 1 gennaio 2000 al 31 dicembre 2020 (252 osservazioni).

3 Analisi Preliminare

Una prima esplorazione grafica dei dati, mostra una stagionalità annuale abbastanza evidente e un trend non regolare: si ha infatti un andamento costante fino al 2008, dove si registra un picco negativo delle vendite dovuto alla crisi globale conseguente allo scoppio della bolla immobiliare americana;

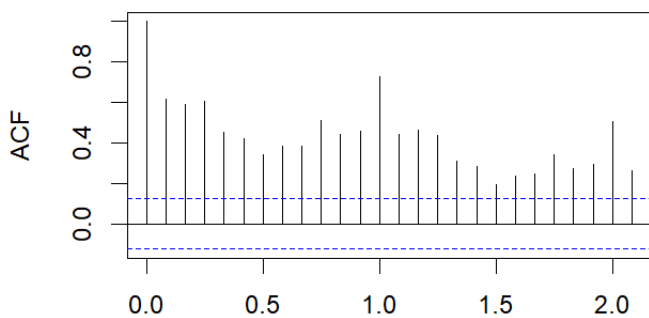
dal 2008 al 2014 si registra invece una crescita che riporta il trend ai livelli pre-2008, e che saranno mantenuti fino al 2020, dove viene registrato un altro picco negativo, questa volta dovuto alla pandemia di Covid-19.



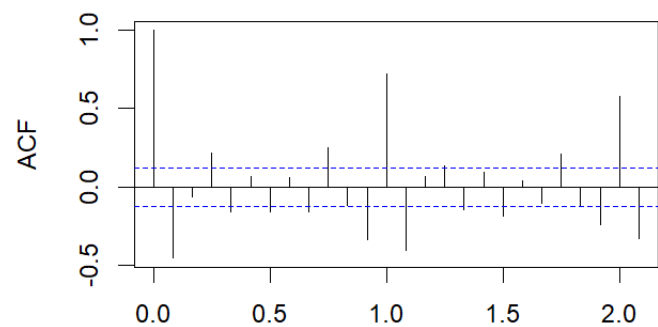
Dati i picchi positivi nei grafici su due periodi della funzione di autocorrelazione e della funzione di autocorrelazione al netto del trend, si può concludere che la serie storica presenti un andamento stagionale annuale.

Inoltre, andando a confrontare la stagionalità su anni differenti delle vendite normalizzate, è possibile notare che le stagionalità più "rosse", corrispondenti agli anni meno recenti, abbiano andamenti diversi, soprattutto nei mesi finali dell'anno, rispetto alle stagionalità più chiare, corrispondenti agli anni più recenti. Pertanto la serie potrebbe presentare una stagionalità non uniforme.

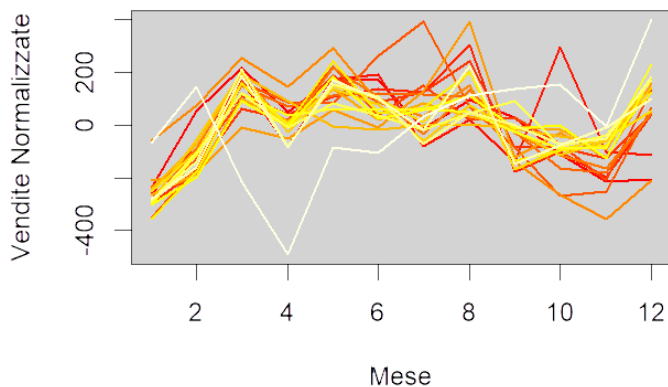
Series data_ts



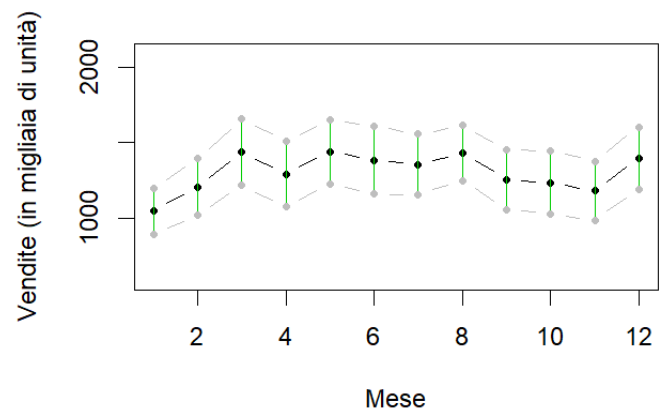
Series diff(data_ts)



Confronto Stagionalità



Bande di Confidenza della Stagionalità

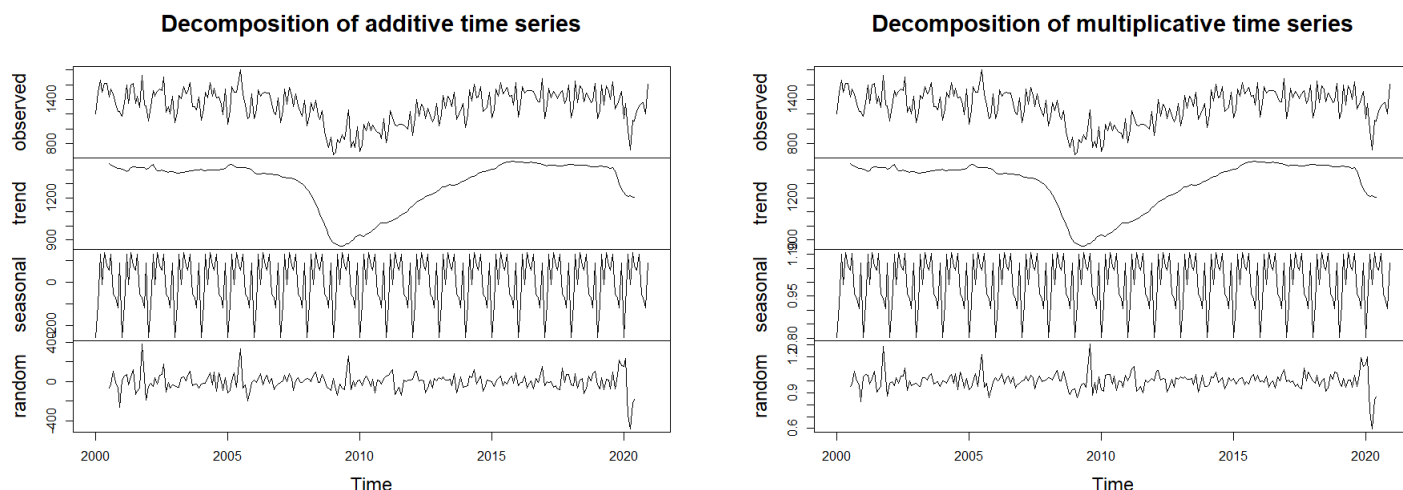


4 Decomposizione

4.1 Decomposizione Stagionalità Stazionaria

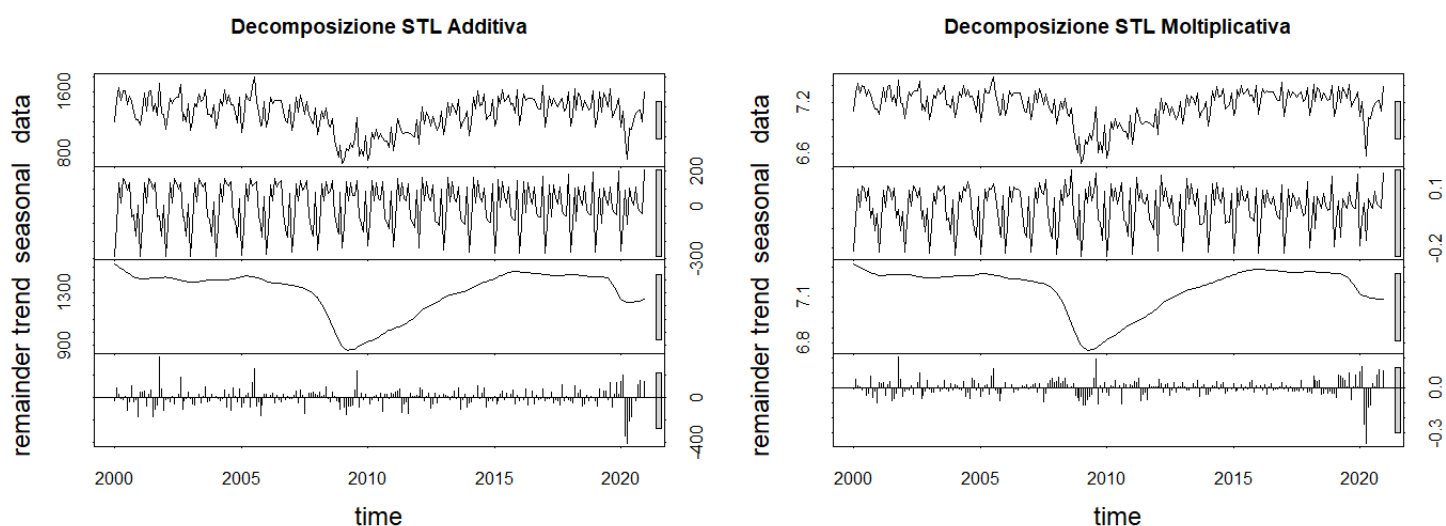
Il modello di decomposizione additiva e il modello di decomposizione moltiplicativa restituiscono risultati molto simili, sembra però che il modello additivo sia in grado di filtrare meglio il rumore dovuto al picco della crisi del 2008.

Troviamo un'ulteriore conferma della presenza (e della rilevanza) di stagionalità nella serie notando che la stagionalità ha un ordine di grandezza maggiore di quello del rumore.



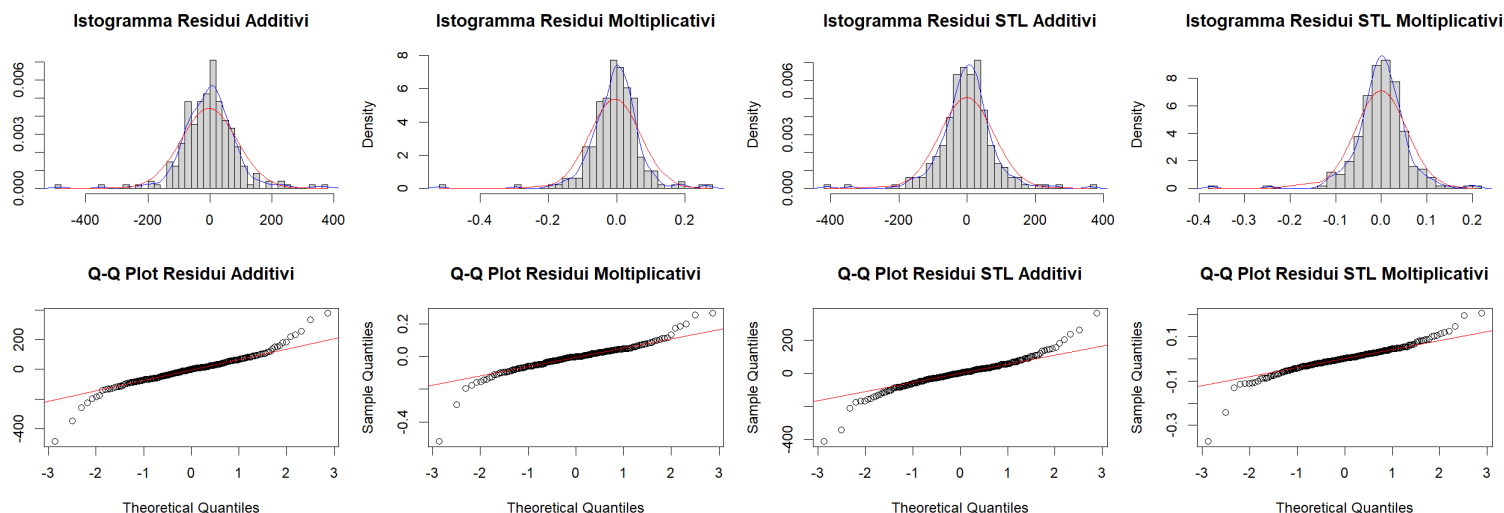
4.2 Decomposizione STL

Come precedentemente accennato nell'analisi preliminare, la serie potrebbe presentare una stagionalità non uniforme e pertanto è stato ritenuto necessario utilizzare anche metodi di decomposizione STL. Anche con questi modelli si nota come il modello additivo presenti picchi di rumore più smorzati rispetto a quello moltiplicativo.

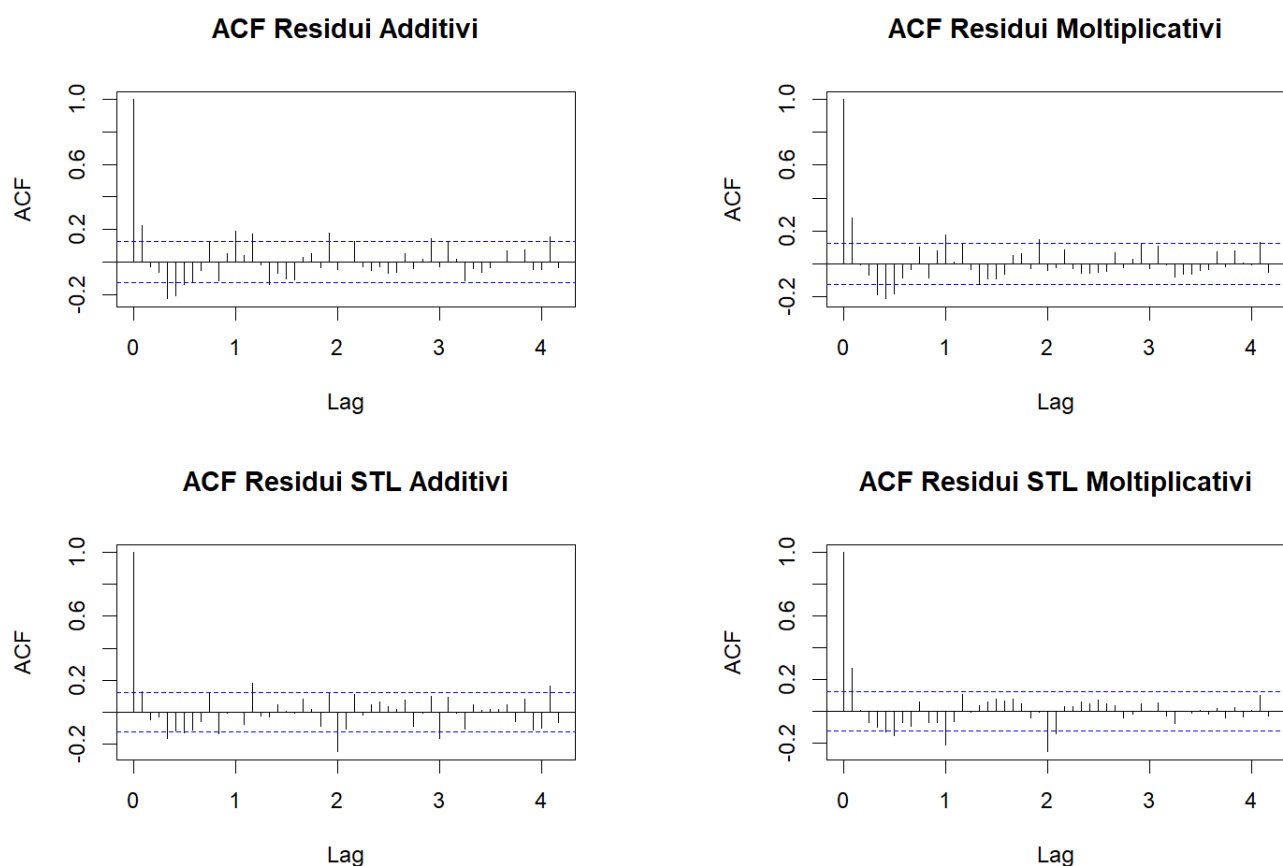


4.3 Analisi dei Residui

L'analisi dei residui restituisce risultati simili per tutti e 4 i modelli analizzati: i residui hanno una distribuzione casuale e non sembrano aver catturato stagionalità; in nessun caso, però, risultano essere Gaussiani dato che i test di Shapiro Wilk danno sempre pessimi p-value, principalmente per la presenza di residui nelle code della gaussiana.



Osservando i grafici delle autocorrelazioni, si nota come i residui dei modelli STL siano in grado di filtrare maggiormente il picco negativo iniziale e come, in generale, catturino meno struttura rispetto ai due modelli non STL.



Per scegliere quale dei due modelli STL utilizzare, è stato effettuato un confronto tra le deviazioni standard della funzione di autocorrelazione (per completezza sono stati riportati i risultati anche dei modelli additivo e moltiplicativo con stagionalità stazionaria). Per le considerazioni fatte in precedenza e per i risultati del confronto delle deviazioni standard, la scelta è ricaduta sul modello di **decomposizione STL additivo**.

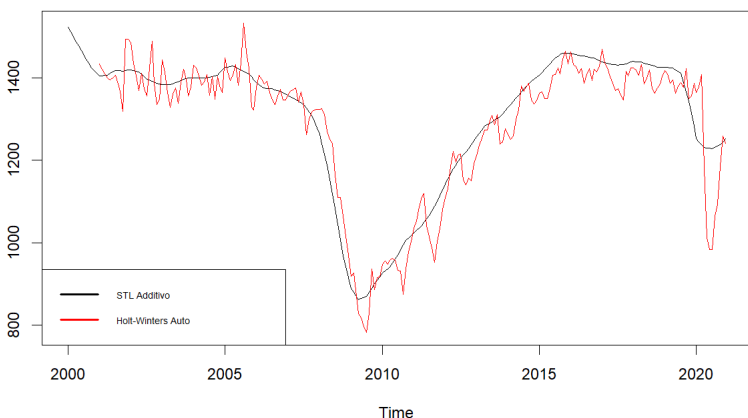
```
> sd(acf(data.ts.dar, plot=F)$acf) # Modello di Decomposizione Additivo
[1] 0.2426831
> sd(acf(data.ts.dmr, plot=F)$acf) # Modello di Decomposizione Moltiplicativo
[1] 0.2395744
> sd(acf(data.ts.stlar, plot=F)$acf) # Modello di Decomposizione STL Additivo
[1] 0.2282995
> sd(acf(data.ts.stlmr, plot=F)$acf) # Modello di Decomposizione STL Moltiplicativo
[1] 0.2324739
```

5 Previsione

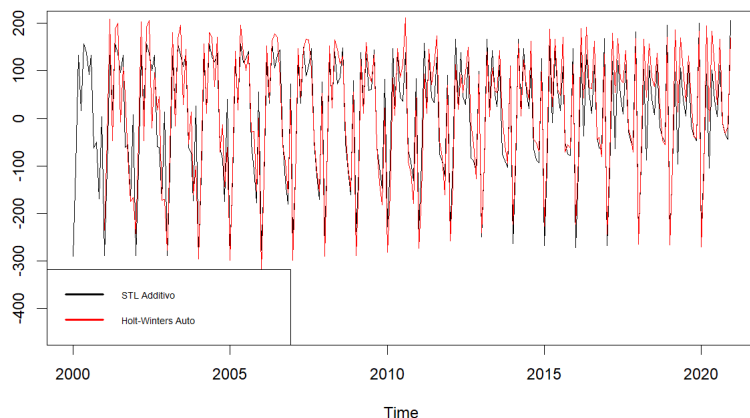
5.1 Holt Winters

Il primo modello di Holt Winters considerato possiede i seguenti parametri, calcolati automaticamente: $\alpha = 0.36, \beta = 0, \gamma = 0.40$. Di seguito i grafici di confronto con i risultati della decomposizione.

Confronto HW Auto con Trend STL Additivo



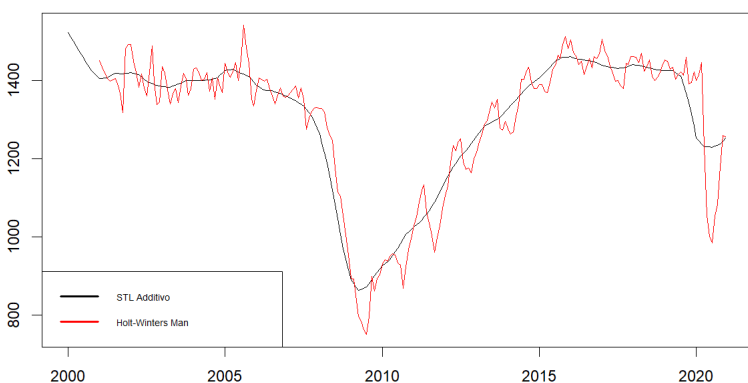
Confronto HW Auto con Stagionalità STL Additivo



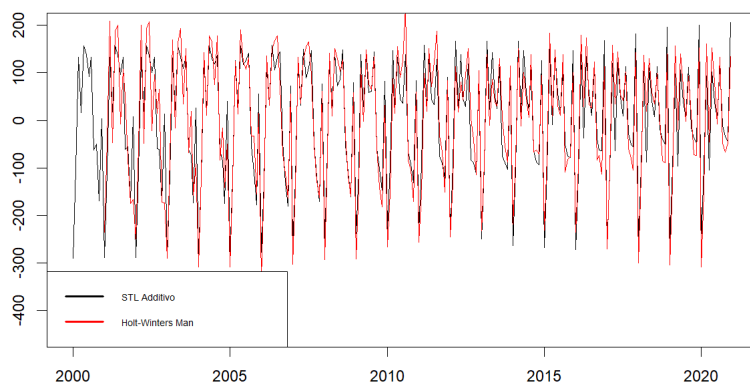
È stato valutato anche un modello Holt Winters con i parametri $\alpha = 0.35, \beta = 0.1, \gamma = 0.45$ settati manualmente e con le condizioni iniziali settate mediante regressione lineare, con l'obiettivo di ottenere un andamento più aderente a quello del trend STL, soprattutto per i valori successivi al 2011. In entrambi i modelli di Holt Winters considerati, la stagionalità risulta avere lo stesso periodo di quella ottenuta con la decomposizione ma un'ampiezza, in media, leggermente superiore.

Osservando i grafici delle previsioni riportati di seguito, si nota come entrambi i modelli riportino un trend in crescita nel prossimo futuro, il modello settato manualmente risulta però più ottimista. In verde sono rappresentati gli intervalli di confidenza non parametrici (l'analisi dei residui metterà in evidenza la non gaussianità di quest'ultimi).

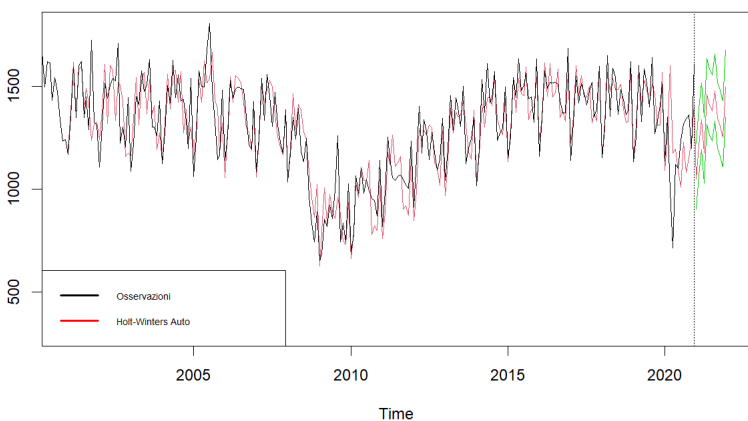
Confronto HW Man con Trend STL Additivo



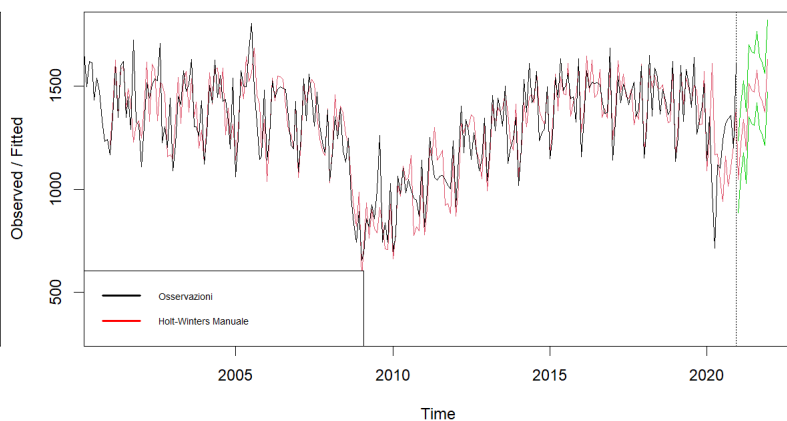
Confronto HW Man con Stagionalità STL Additivo



Holt-Winters Auto

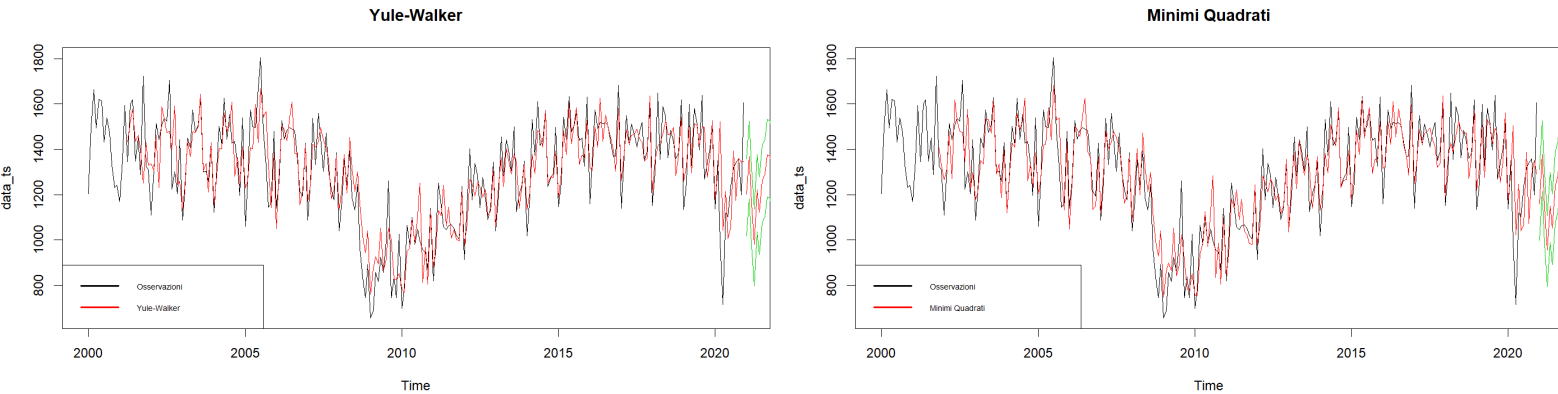


Holt-Winters Manuale



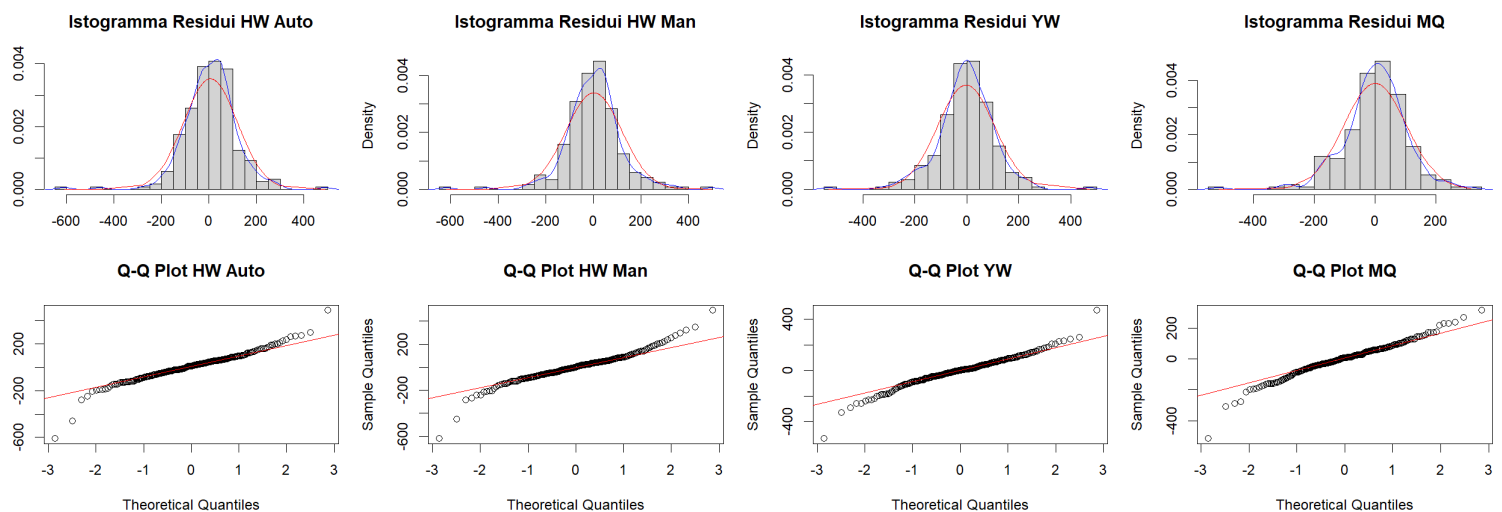
5.2 Metodi Autoregressivi

Son stati considerati per la previsione anche due modelli autoregressivi: *Yule-Walker* e *Minimi Quadrati*. Come visibile dai seguenti grafici, i metodi restituiscono previsioni che differiscono da quelle ottenute con il metodo Holt-Winters: con questi modelli, le previsioni sembrano essere meno ottimiste suggerendo un andamento più stazionario del trend. In verde sono riportati gli intervalli di confidenza delle previsioni che, anche questa volta, sono non parametrici per analoghe considerazioni.

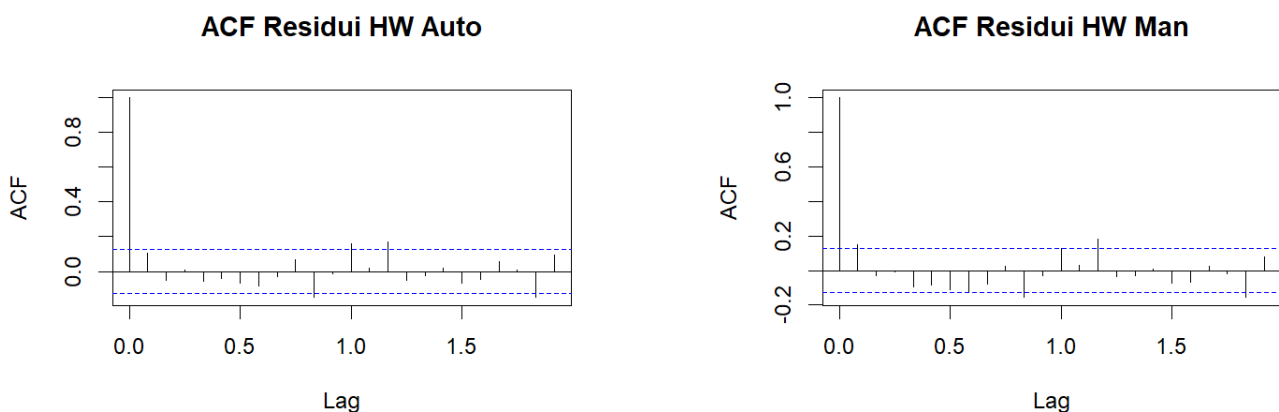


5.3 Analisi dei Residui

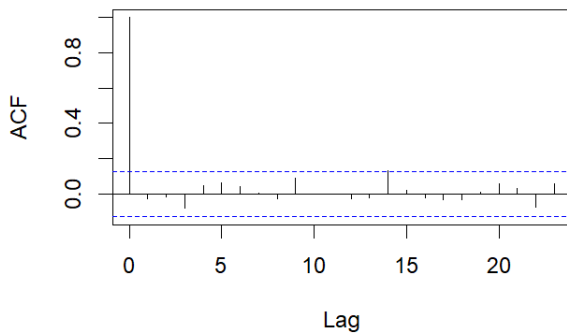
L'analisi dei residui dei metodi introdotti in precedenza, restituisce risultati molto simili in tutti i casi: tutti i modelli hanno residui con distribuzioni casuali non gaussiane dato che in tutti i casi il p-value del test di Shapiro-Wilk risulta essere molto ridotto.



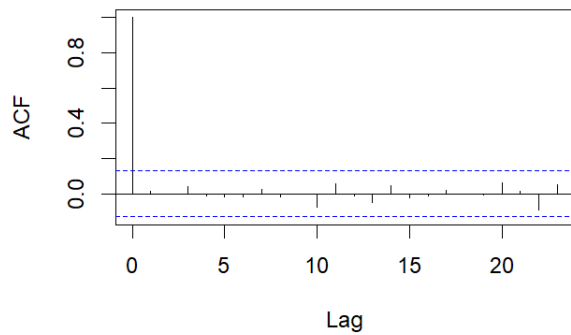
Come visibile dai grafici delle funzioni di autocorrelazione dei residui, i residui dei due modelli di Holt-Winters analizzati sembrano possedere una piccola porzione di struttura residua a differenza dei residui dei due modelli auto regressivi.



ACF Residui YW



ACF Residui MQ



5.4 Confronto

Per testare la capacità predittiva dei metodi analizzati, è stata effettuata un'autovalidazione sui dati dell'ultimo anno della serie, ovvero il 2020. I risultati ottenuti non sono stati dei migliori: l'errore medio inferiore è stato ottenuto con il metodo Yule-Walker, ma comunque siamo lontani da risultati soddisfacenti.

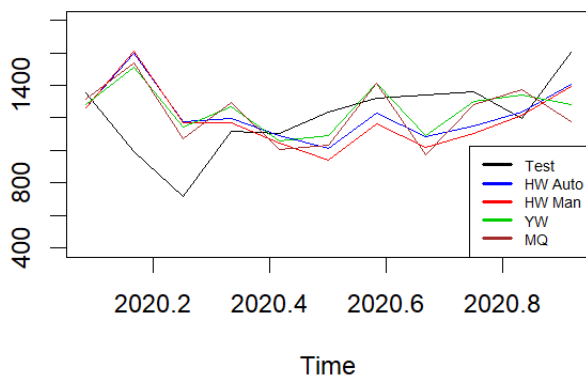
```
> sqrt(mean((test-res.hw.1)^2)) # Holt-Winters con parametri automatici
[1] 270.4506
> sqrt(mean((test-res.hw.2)^2)) # Holt-Winters con parametri manuali
[1] 290.6424
> sqrt(mean((test-res.ar)^2)) # Yule Walker
[1] 252.7065
> sqrt(mean((test-res.ls)^2)) # Minimi Quadrati
[1] 281.4944
```

L'accuratezza dei metodi valutata solo su dati relativi all'annata 2020 non è sufficientemente affidabile in quanto fortemente influenzata da fattori esterni che hanno indotto comportamenti anomali nei dati. Pertanto è stato ritenuto necessario ripetere il confronto considerando i dati del 2019 come test set e allenando i modelli con i dati fino a Dicembre 2018.

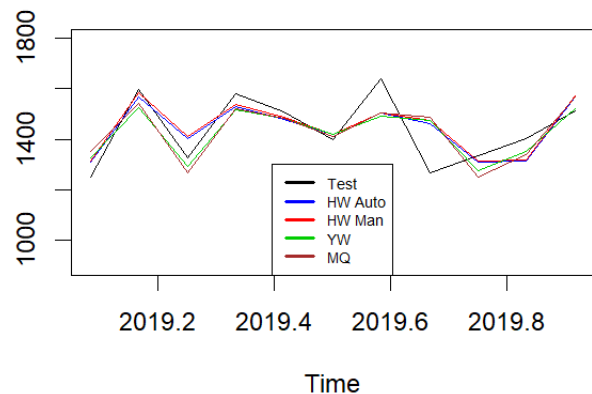
Questa volta i risultati sono stati decisamente migliori: il modell di Holt-Winters automatico risulta essere il migliore, seguono il modello di Holt-Winters manuale e il modello di Yule Walker che hanno errori medi molto simili.

```
> sqrt(mean((test-res.hw.1)^2)) # Holt-Winters con parametri automatici
[1] 86.16626
> sqrt(mean((test-res.hw.2)^2)) # Holt-Winters con parametri manuali
[1] 88.69192
> sqrt(mean((test-res.ar)^2)) # Yule Walker
[1] 89.98988
> sqrt(mean((test-res.ls)^2)) # Minimi Quadrati
[1] 95.88027
```

Confronto 2020



Confronto 2019



6 Conclusioni

Per effettuare la previsione finale è stato preferito il metodo di Yule Walker: i residui del metodo non presentano struttura residua, ha ottenuto i migliori risultati nell'autovalidazione con i dati del 2020, e buoni risultati nell'autovalidazione con i dati del 2019.

Il metodo prevede una crescita nella vendita di autoveicoli nell'anno 2021; per quanto riguarda la stagionalità, è presente un netto calo di circa 600-700 migliaia di unità nel mese di febbraio mentre, dalla metà del mese di marzo, è prevista una crescita più o meno costante fino alla fine dell'anno.

