

Progetto III

Simone Manti, matricola: 566908

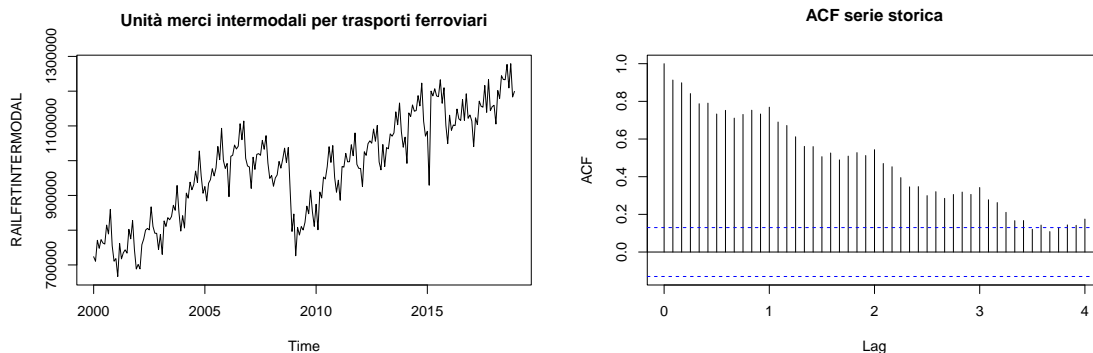
28/12/2021

Introduzione

La seguente analisi è rivolta alle aziende dedite alla costruzione di trasporti ferroviari per merci intermodali. Analizzeremo una serie storica (dettagli in fondo alla relazione) riguardante il traffico ferroviario di merci intermodali e il nostro obiettivo sarà predire il numero di unità nel 2019, in modo da costruire mezzi di trasporto che siano sufficientemente capienti ed efficienti.

Presentazione del problema

Cominciamo l'analisi osservando una rappresentazione grafica della serie storica e la sua funzione di autocorrelazione.

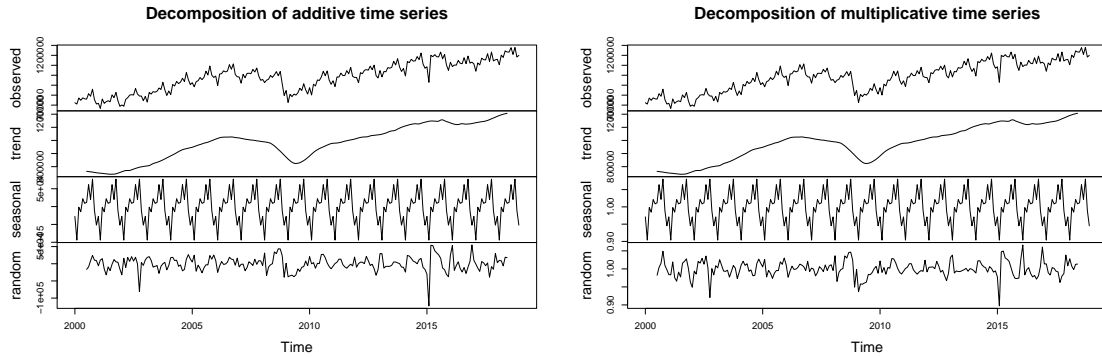


Prima di continuare, diamo alcuni chiarimenti sulla serie storica e facciamo alcune osservazioni.

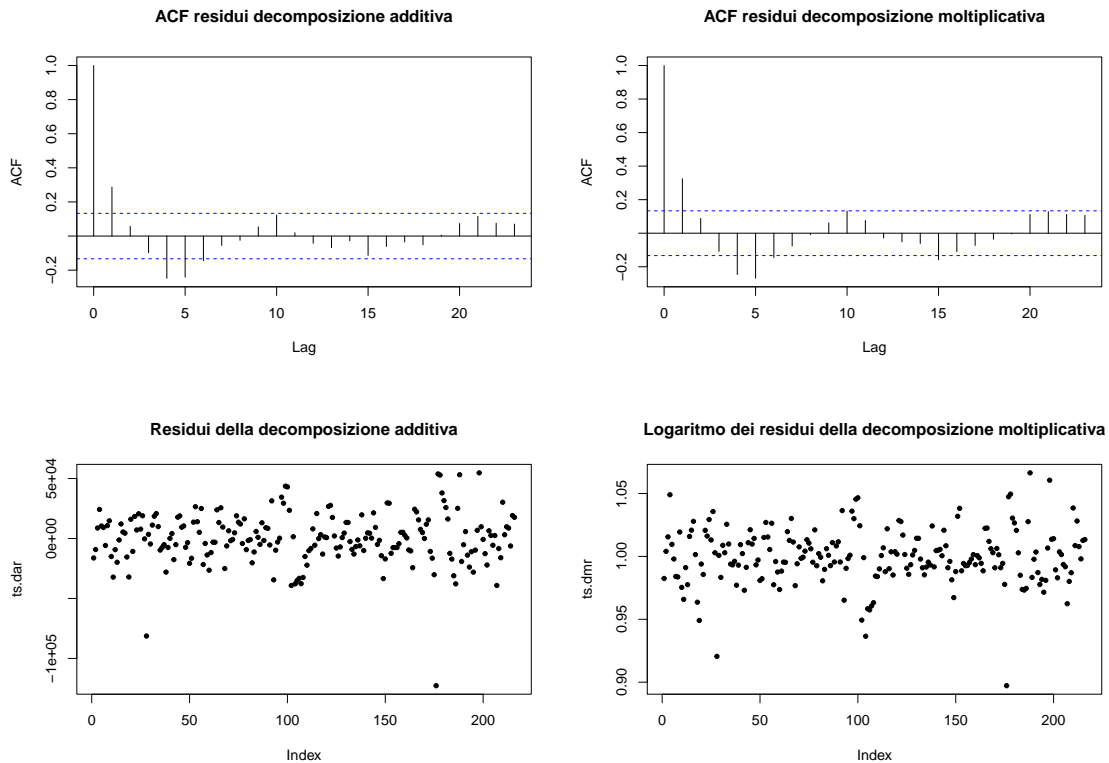
1. La serie storica rappresenta il numero di unità di merci intermodali trasportate per via ferroviaria per ogni mese, dal Gennaio 2000 al Dicembre 2018.
2. Guardando l'autocorrelazione osserviamo immediatamente la presenza di un trend e anche di stagionalità, data la presenza di "picchi" in ogni periodo.

Decomposizione

Procediamo la nostra analisi guardando come si comportano i metodi di decomposizione additivo e moltiplicativo con stagionalità fissa.

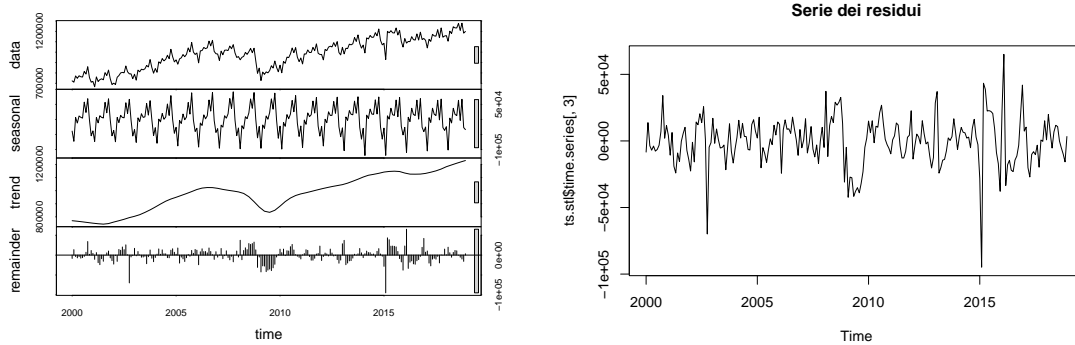


Osserviamo che il metodo di decomposizione additivo sembra catturare meglio la componente stagionale, in quanto le ordinate variano in $[10^{-5}, 10^5]$. Anche i residui però variano nello stesso intervallo, proviamo a guardare la loro autocorrelazione e il loro plot per ambedue i metodi.



Effettivamente i residui hanno un ordine di grandezza pari a 10^4 , solo pochi sono dell'ordine di 10^5 . Quindi sia nella decomposizione additiva che in quella moltiplicativa non riusciamo a catturare sufficientemente bene la stagionalità, in quanto in ambedue i casi gli ordini di grandezza sono comparabili (in questo caso addirittura gli stessi).

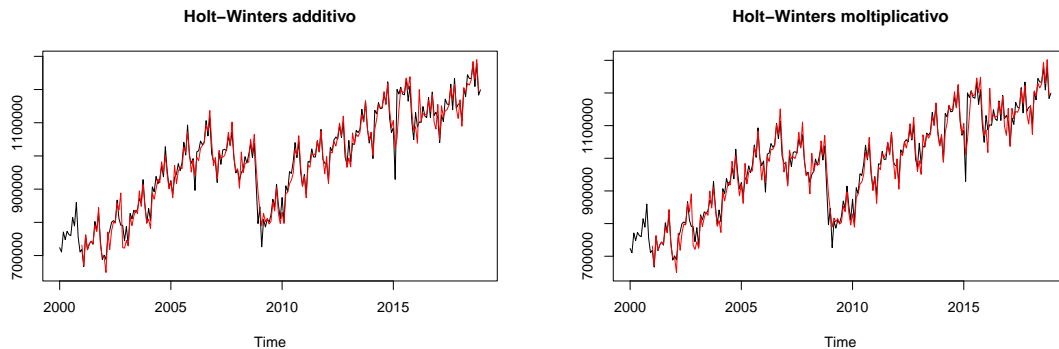
Dato che entrambe le decomposizioni a stagionalità fissa non restituiscono risultati convincenti, proviamo a decomporre la serie storica con il metodo a stagionalità variabile.



Ci troviamo sostanzialmente nella stessa situazione di prima, anche in questo caso non abbiamo un risultato convincente, nel senso che non riusciamo a catturare sufficientemente bene la stagionalità. Non crediamo che la stagionalità sia trascurabile visto anche l'aspetto della funzione di autocorrelazione della serie.

Holt Winters

Consideriamo adesso il metodo di Holt-Winters e confrontiamo il metodo di HW additivo e il metodo moltiplicativo, per capire quale dei due funziona meglio per la serie storica in esame. Riportiamo in seguito una rappresentazione grafica dei due metodi con coefficienti di default (i.e. i coefficienti ottimali ottenuti minimizzando la varianza dei residui) e i valori di tali coefficienti.



	HW Additivo	HW Moltiplicativo
alpha	0.672407133208956	0.61729055006991
beta	0.0114690110333158	0.00945264949245809
gamma	0.425384965516173	0.491368244581332

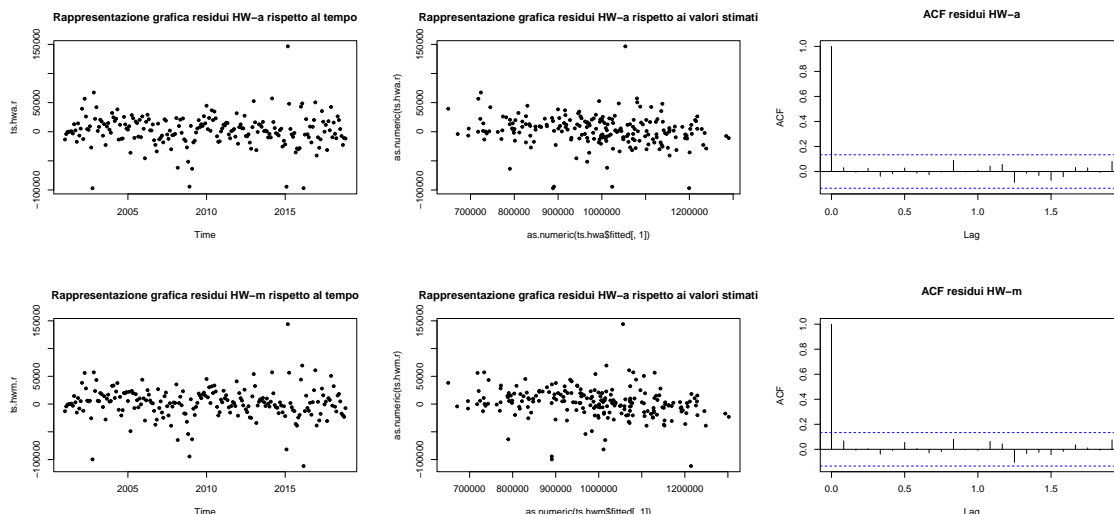
Facciamo alcune osservazioni sui coefficienti.

- Il valore di alpha è molto grande per entrambi i metodi (circa 0.6) e quindi ambedue i metodi prediligono la componente innovativa: infatti questo si vede anche dai plot delle due serie generate, entrambe seguono molto la serie originale.
- Il valore di beta è molto basso per entrambi i metodi (circa 0.01): questo è molto ragionevole, data la forma che ha il trend, vista nella sezione Decomposizione ("oscilla poco").
- Il valore di gamma è molto vicino a 0.5 per entrambi i metodi (più per il metodo moltiplicativo).

Guardando i plot e le osservazioni fatte poc'anzi entrambi i metodi (sia additivo che moltiplicativo) sembrano catturare già bene la struttura della serie storica in esame, soprattutto nella parte finale (che è la più interessante per la predizione). Confrontiamo adesso i due metodi analizzandone in dettaglio i residui

Varianza non spiegata HW additivo: 0.03298328

Varianza non spiegata HW moltiplicativo: 0.03533351

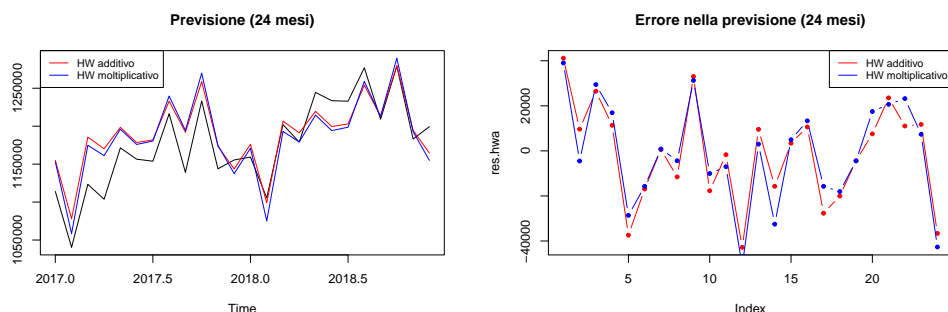


Entrambi i metodi hanno una struttura molto simile dei residui. Inoltre, guardando le funzioni di autocorrelazione osserviamo che entrambi i metodi riescono a catturare sufficientemente bene le informazioni su trend e stagionalità della serie, poichè per ogni lag (eccetto, ovviamente, $\text{lag} = 0$) la ACF non è significativamente diversa da 0. Ambedue presentano una proporzione di varianza non spiegata molto bassa (leggerissimamente meglio HW-a).

Procediamo la nostra analisi dei due metodi di HW confrontando le loro capacità predittive: per farlo valutiamo l'errore nella predizione per 2 anni tramite un procedimento di autovalutazione.

Varianza dell'errore nel metodo additivo: 21920.17

Varianza dell'errore nel metodo moltiplicativo: 22854.78



Possiamo fare alcune osservazioni.

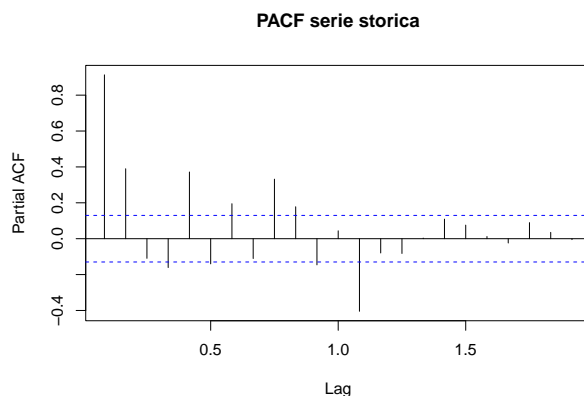
1. Negli ultimi due anni i valori della serie sono dell'ordine di 10^6 e per entrambi i metodi abbiamo una varianza dell'errore pari a circa 2×10^4 , quindi entrambi i metodi hanno un'incertezza sulla previsione pari a circa il 2
2. Tra i due preferiamo il metodo di HW additivo, sia perchè ha una varianza dell'errore leggermente più bassa sia perchè il metodo di HW moltiplicativo ha problemi di convergenza nel dodicesimo mese di

validazione (come si vede facilmente dal plot).

Concludiamo dalla nostra analisi che il metodo di Holt-Winters additivo si comporta meglio di quello moltiplicativo.

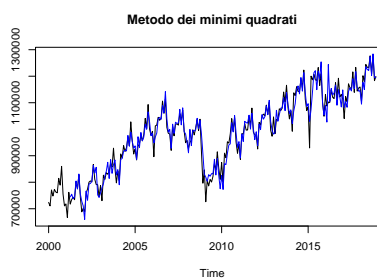
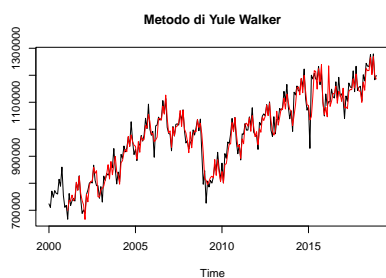
Metodi autoregressivi

Procediamo la nostra analisi confrontando due metodi autoregressivi: il metodo di Yule-Walker ed il metodo autoregressivo dei minimi quadrati. Prima di farlo, guardiamo la funzione di autocorrelazione parziale della serie storica in esame.



Dal plot della PACF visto poc'anzi ci aspettiamo che i metodi autoregressivi dipendano dai primi 13 istanti. Riportiamo adesso una rappresentazione grafica dei due metodi e una tabella con i valori dei coefficienti scelti.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15
YW	0.77	0.27	-0.05	-0.23	0.18	-0.02	0.07	-0.26	0.12	0.20	-0.04	0.35	-0.40	NA	NA
OLS	0.70	0.20	0.08	-0.25	0.20	0.05	0.02	-0.26	0.15	0.21	-0.08	0.48	-0.38	0.06	-0.19

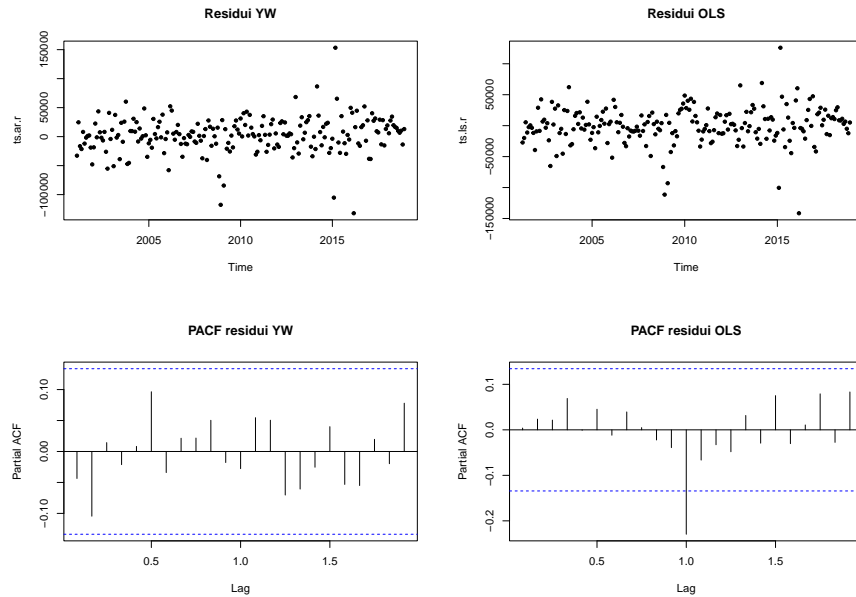


Come ci aspettavamo, il metodo di YW utilizza 13 coefficienti, mentre inaspettatamente il metodo dei minimi quadrati ne utilizza 15. Osserviamo immediatamente dai plot che il metodo di YW sembra molto meno preciso (verso la fine della serie) rispetto al metodo dei minimi quadrati, che invece sembra catturare meglio la struttura della serie.

Analizziamo anche in questo caso i residui per ambedue i metodi.

Varianza non spiegata Yule Walker: 0.04431582

Varianza non spiegata Ordinary Least Squares: 0.04043381

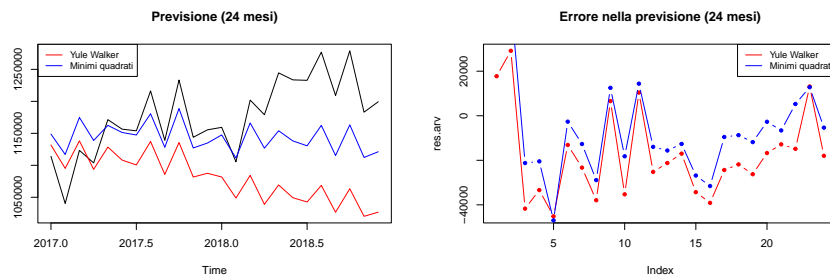


La struttura dei residui per entrambi i metodi sembra molto simile a quella dei metodi HW, anche se la PACF del metodo dei minimi quadrati presenta un picco in corrispondenza del periodo.

Confrontiamo adesso le due capacità predittive tramite un autovalidazione.

Varianza dell'errore nel metodo Yule Walker: 26305.48

Varianza dell'errore nel metodo dei minimi quadrati: 21972.99



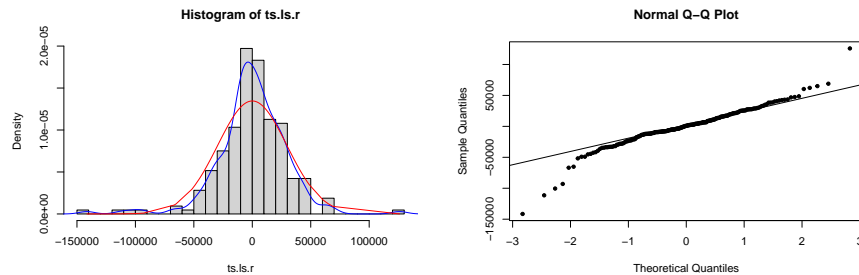
Dai grafici precedenti osserviamo che non c'è paragone: Yule Walker si discosta troppo dalla serie originale e ha una varianza dell'errore decisamente più alta di quella dell'OLS. Certamente il miglior metodo tra i due è l'OLS.

Scelta finale del metodo e previsione

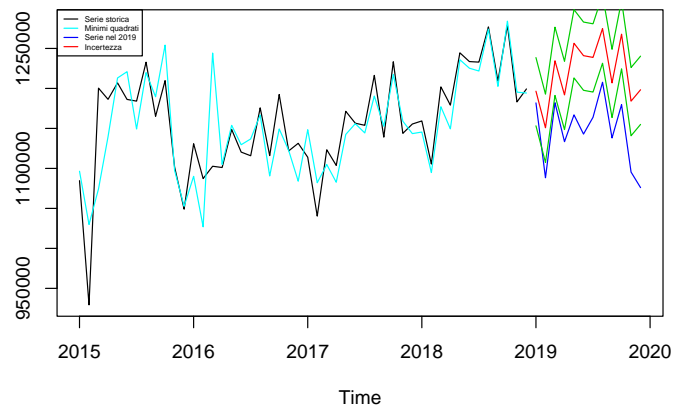
Scegliamo, infine, il miglior metodo tra quelli esaminati per poi procedere alla previsione del numero di unità di merci intermodali per il 2019. Il miglior metodo è il metodo dei minimi quadrati: anche se a discapito di Holt Winters additivo ha una varianza dell'errore leggermente più grande riesce a catturare molto meglio la parte finale della serie. Infatti, questo si può vedere sia dai grafici degli errori nelle previsioni (l'errore nei mesi finali di test è molto più piccolo per OLS rispetto a HW additivo), sia dai grafici delle previsioni (la serie delle previsioni cresce leggermente per OLS nei mesi finali, mentre per HW additivo continua a scendere). Calcoliamo l'incertezza per il metodo OLS: prima di farlo, analizziamo i residui per decidere se rigettare o meno l'ipotesi di gaussianità.

##

```
## Shapiro-Wilk normality test
##
## data: ts.ls.r
## W = 0.93539, p-value = 4.236e-08
```



La distribuzione empirica dei residui (in blu) si discosta tanto dalla distribuzione teorica (in rosso). Inoltre, il test di SW restituisce un p-value molto basso, dunque rigettiamo decisamente l'ipotesi gaussiana. Possiamo, quindi, utilizzare solo incertezze empiriche. Procediamo allora con la previsione del numero di unità nel 2019. Fortunatamente, dato che siamo nel 2021, possiamo confrontare la nostra previsione con i veri risultati.



Il risultato della previsione è in linea con la stima dell'errore fatta prima, inoltre l'andamento della previsione è molto simile all'andamento effettivo della serie.

Conclusione

Per concludere, abbiamo confrontato diversi metodi di analisi di serie storiche per studiare la serie storica riguardante il numero di unità di merci intermodali per trasporti ferroviari. Fra tutti, il migliore è risultato il metodo autoregressivo dei minimi quadrati, il quale si è rivelato decisamente notevole in fase di previsione.