

Progetto di Statistica II

Luca Ferragina

23 dicembre 2016

Indice

1	Valore della moneta italiana nella storia	2
1.1	Introduzione	2
1.2	Grafico e autocorrelazione	2
1.3	Decomposizione della serie	3
1.4	Previsione della serie	5
1.5	Analisi dei residui	5
1.6	Bande di confidenza per la previsione	7
1.7	Serie troncata	8
1.8	Autovalidazione	10
2	Temperature nello stato dell'Indiana	11
2.1	Grafico e autocorrelazione	11
2.2	Decomposizione della serie	12
2.3	Analisi e previsione del trend	13
2.4	Previsione della serie	16
2.5	Analisi dei residui	18
2.6	Bande di confidenza della previsione	20

Capitolo 1

Valore della moneta italiana nella storia

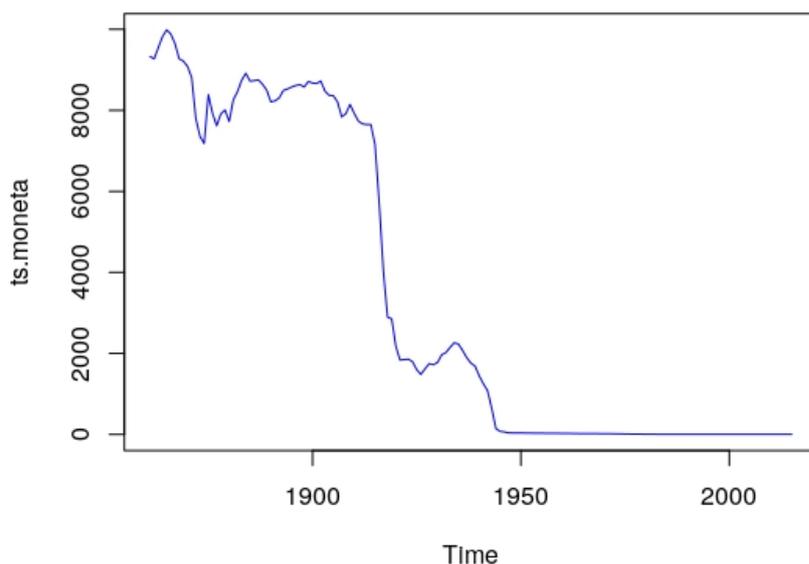
1.1 Introduzione

«Le trasformazioni economiche verificatesi nel nostro Paese nel corso della sua storia e il fenomeno delle variazioni temporali del valore della moneta determinano la necessità di disporre di uno strumento statistico che consenta confronti omogenei, dal punto di vista del potere di acquisto della moneta, tra valori riferiti a epoche diverse.»

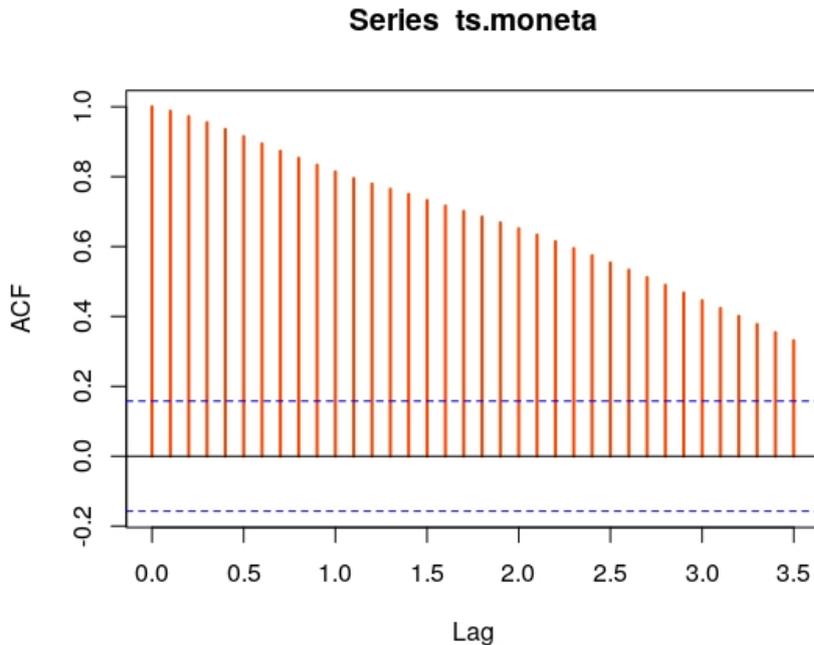
Queste sono le parole di apertura del documento con cui l'Istat accompagna la pubblicazione della serie storica dell'indice nazionale dei prezzi di consumo per le famiglie di operai e impiegati. Tale strumento consiste nella media (ponderata opportunamente) dei prezzi di particolari aggregati di beni e servizi che rappresentano i consumi di una famiglia la cui persona di riferimento è un lavoratore dipendente. Esso è un buon indicatore delle variazioni del potere di acquisto della moneta nel corso del tempo, tanto che fino al 1967 veniva denominato indice del costo della vita e viene tuttora utilizzato per scopi legali, ad esempio l'adeguamento degli affitti o degli assegni di mantenimento. L'indice è registrato a partire dall'anno 1861, ha cadenza annuale ed è normalizzato rispetto all'anno 2015 nel senso che il valore di un certo anno indica di quante volte occorre moltiplicare una quantità monetaria riferita a quell'anno per riportarla al valore del 2015.

1.2 Grafico e autocorrelazione

Il grafico della serie storica è il seguente:



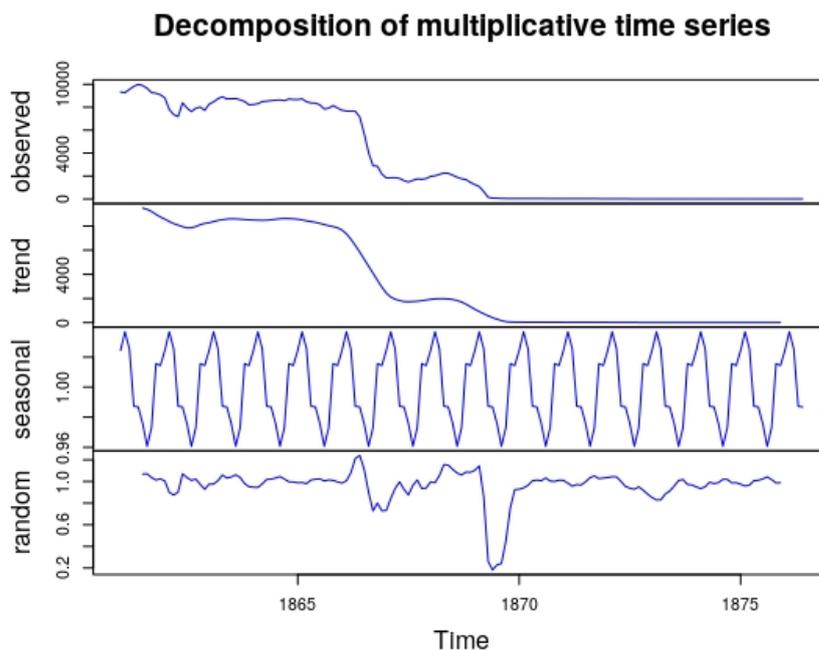
La prima cosa che salta all'occhio è la sostanziale inesistenza di stagionalità a favore di un trend dominante che a partire dal secondo dopoguerra si è fatto sempre più stabile. La funzione di autocorrelazione in effetti è tipica di una serie storica di puro trend:



Osservazione 1.1. Al fine di ottenere un grafico della serie storica in cui l'asse delle x fosse coerente con i dati si è posta la frequenza pari a 1 in via del tutto eccezionale. Come sappiamo, infatti, questa scelta è pericolosa in quanto potrebbe portare ad errori nel calcolo di alcune funzioni (ad esempio il comando `decompose`); nel prosieguo dell'analisi abbiamo posto la frequenza pari a 10 senza alcun motivo particolare, solamente per non incorrere in errori.

1.3 Decomposizione della serie

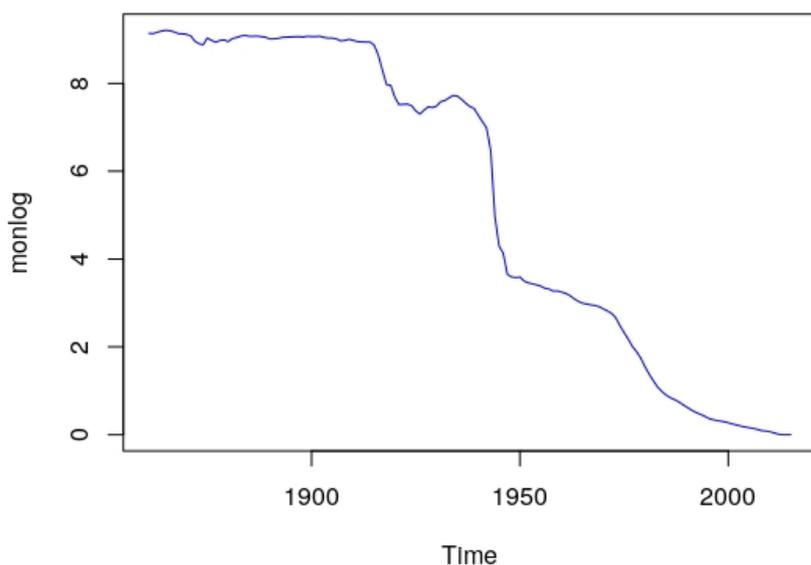
A causa dell'estrema pendenza della serie si è rivelata più ragionevole la decomposizione moltiplicativa rispetto a quella additiva. Qui di seguito ne è riportato il grafico:



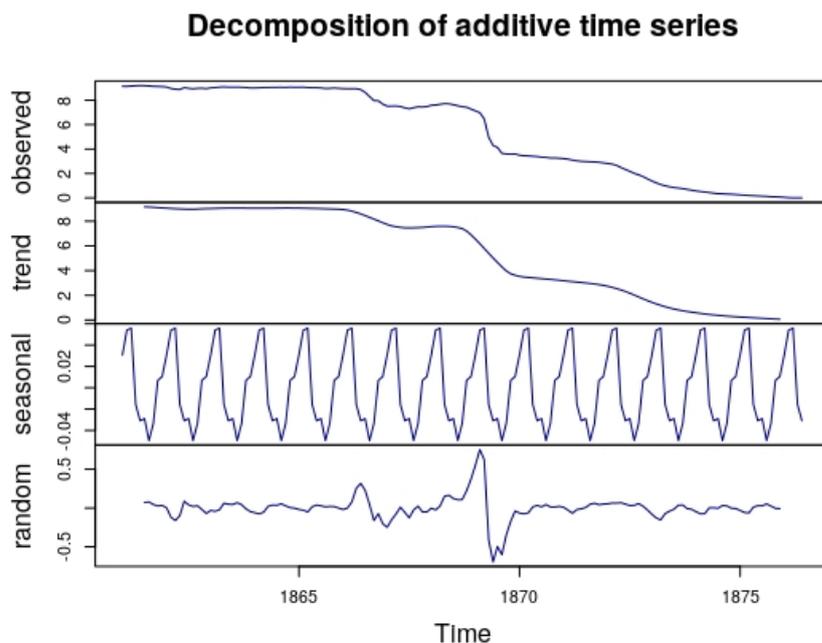
Il trend sembra esser stato catturato in maniera decente, con un andamento molto regolare che accompagna tutta la serie, eccezion fatta per una certa instabilità iniziale (probabile specchio dell'instabilità

economica di un Paese appena unificatosi) e due picchi fortissimamente decrescenti che possiamo associare agli eventi storici delle due guerre mondiali, inoltre pare di scorgere un processo di lenta decrescita avvenuto dal dopoguerra in poi. Per quanto riguarda la stagionalità, possiamo dire che questa è abbastanza contenuta in quanto amplifica il trend di non più del 4%. Infine il rumore è anch'esso generalmente contenuto a parte i soliti due picchi in corrispondenza delle due guerre mondiali.

Serie logaritmica Un altro modo efficace per studiare la natura della serie si è rivelato il passaggio ai dati logaritmici. Il grafico della serie logaritmica è il seguente:



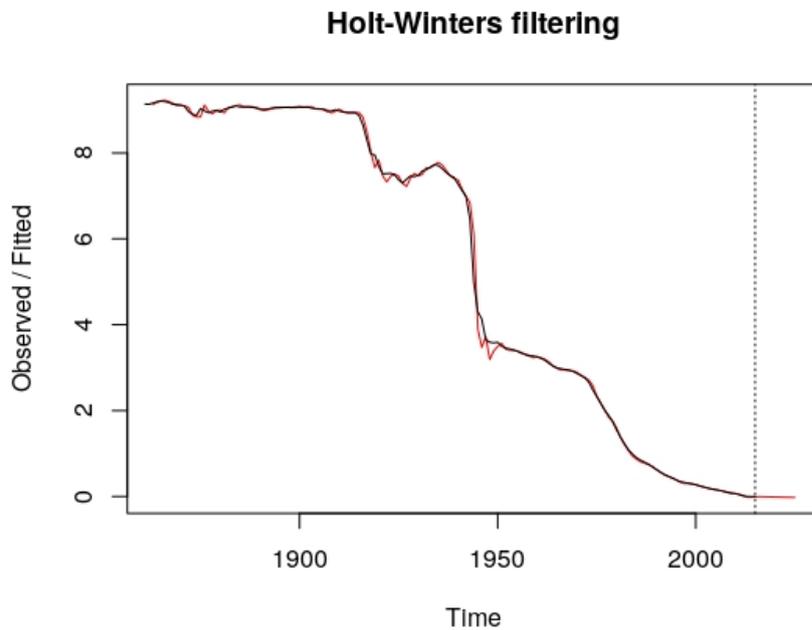
Come si può osservare, grazie ai logaritmi, la scala di valori su cui varia la serie si è ridotta sensibilmente rispetto a quanto avveniva in precedenza, può quindi avere senso ora considerare una decomposizione additiva sulla serie logaritmica:



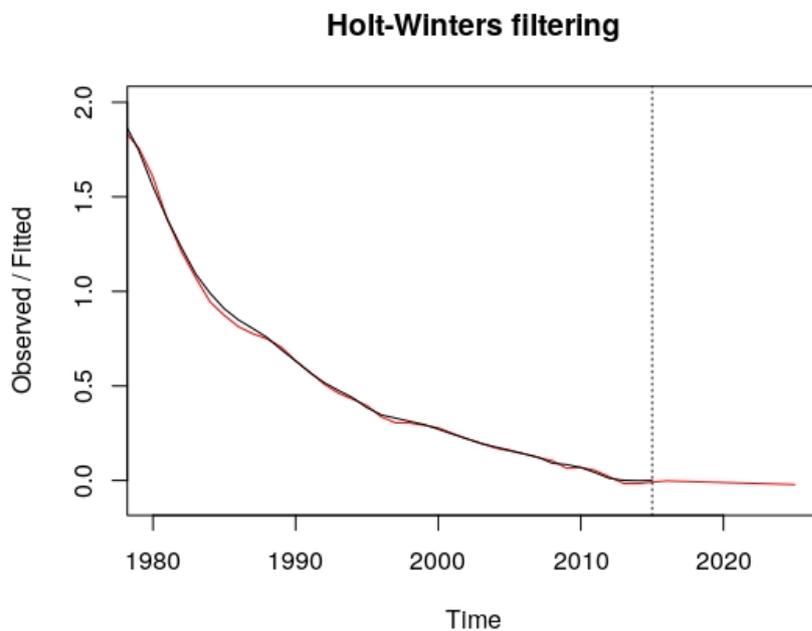
Coerentemente con quanto visto in precedenza si può notare una stagionalità poco significativa e un rumore contenuto con gli stessi picchi di cui abbiamo già parlato. Per il resto dell'analisi abbiamo scelto perciò di focalizzarci sulla serie logaritmica.

1.4 Previsione della serie

La serie (logaritmica) è composta da puro trend che si mantiene piuttosto regolare nella sua parte finale, per cui sembra ragionevole che una buona previsione possa essere ottenuta anche solo tramite il comando SET. Volendo fare una previsione per i prossimi 10 anni, si ottiene il seguente grafico:



Può essere significativo fare lo zoom sugli ultimi anni:

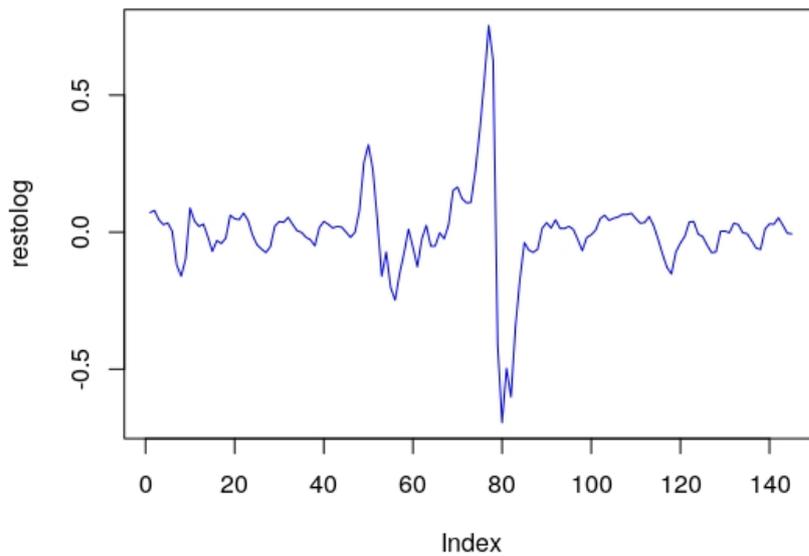


Il valore dell'indice nell'ultimo anno è tornato a crescere (seppur quasi impercettibilmente) dopo molti decenni di lenta discesa, questo fatto tuttavia non ha impedito alla previsione data dal SET di essere pessimistica e di procedere in maniera decrescente.

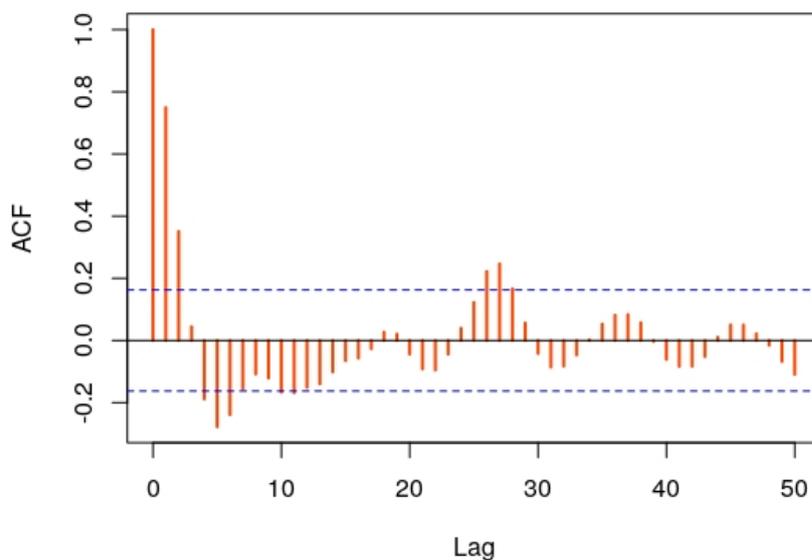
1.5 Analisi dei residui

Sia la decomposizione additiva che il comando SET, i due metodi che abbiamo utilizzato per analisi e previsione della serie logaritmica, dal punto di vista dei residui sono ottimi in quanto hanno varianza spiegata molto vicina ad 1. Anche dal punto di vista della struttura i due residui si somigliano molto, per cui

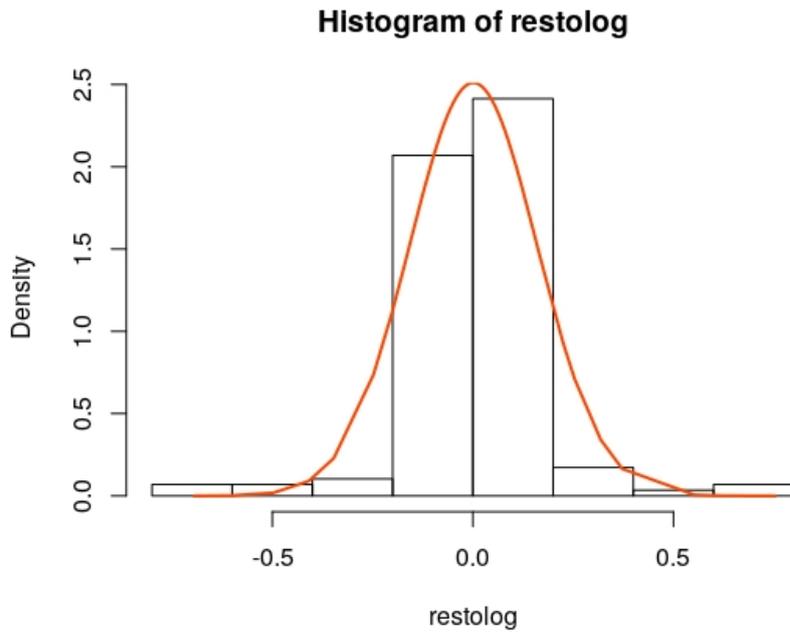
riportiamo solo l'analisi dei residui ottenuti con il metodo SET, di cui qui di seguito mostriamo grafico e funzione di autocorrelazione:



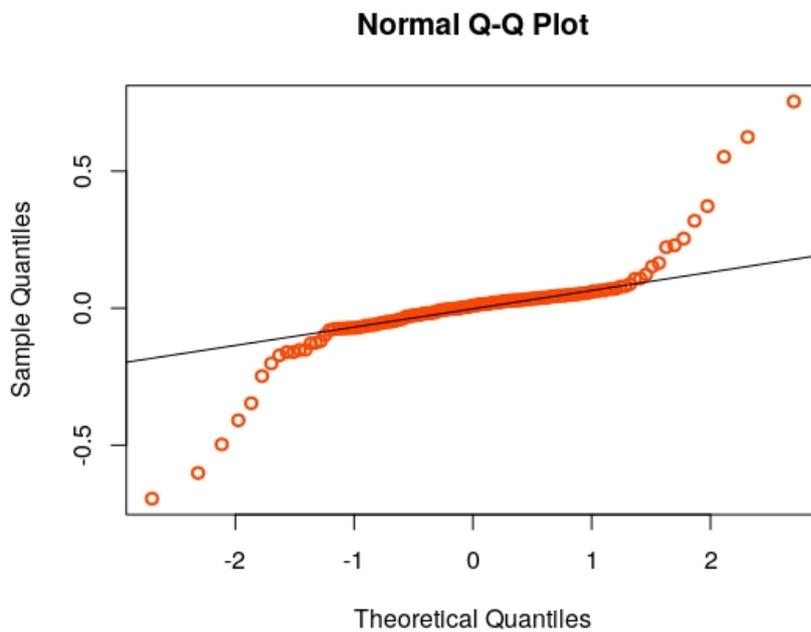
Series restolog



La funzione di autocorrelazione non è ideale ed evidenzia un minimo di struttura all'interno dei residui, tuttavia le linee che superano la banda di confidenza in blu non sono molte, e il fatto che la varianza spiegata sia del 99,8% circa, la più alta dei due metodi, ci permette di andare avanti con l'analisi ignorando questa struttura; inoltre l'istogramma dei residui è molto interessante:



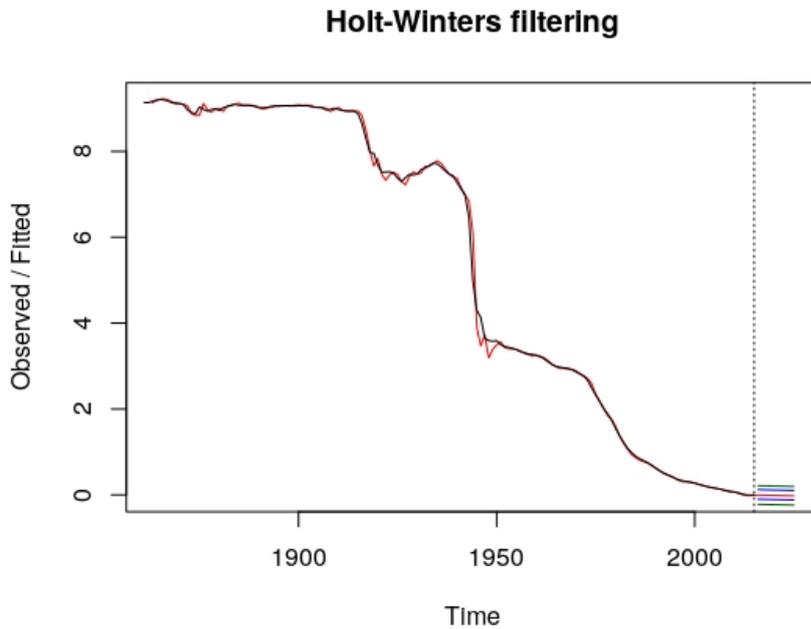
Nonostante una ACF non proprio esaltante l'istogramma è rivelato abbastanza aderente alla densità gaussiana ottenuta con la media e la varianza empiriche. Il *qqplot*, infine, fornisce risultati contrastanti: molto aderenti alla *qqline* per valori interni a -1 e 1 , ma completamente inappropriati per valori esterni.



Quello che si evince dall'analisi è che i residui derivanti dal metodo SET conservano ancora della struttura, come evidenziano ACF e *qqplot*, tuttavia la varianza spiegata è altissima e soprattutto abbiamo riscontrato una buona aderenza fra l'istogramma e la densità gaussiana, quindi nella prossima sezione suporremo comunque, con la dovuta attenzione, la gaussianità dei residui al fine di ottenere bande di confidenza basate su quantili parametrici.

1.6 Bande di confidenza per la previsione

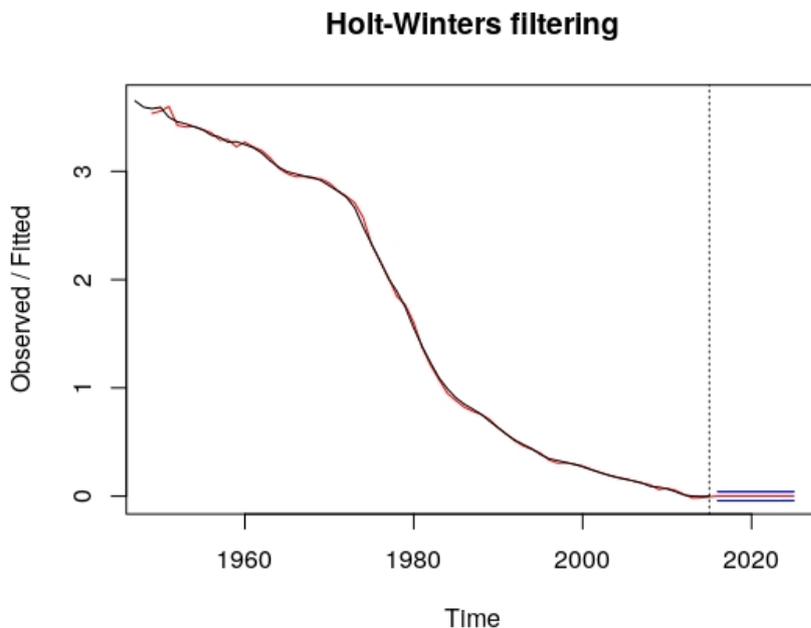
Qui di seguito è esposto il grafico con le bande di confidenza al 90% empirica, in blu, e parametrica di distribuzione gaussiana.



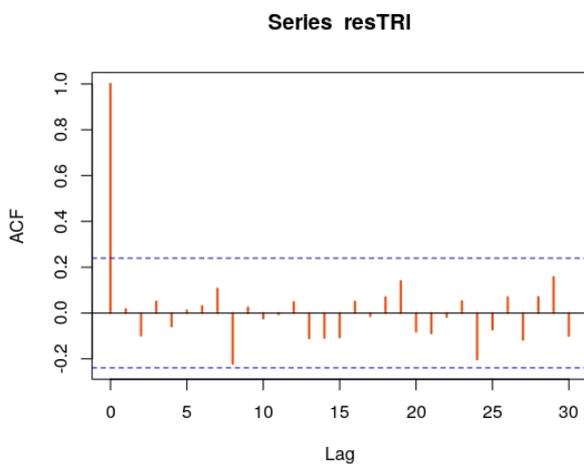
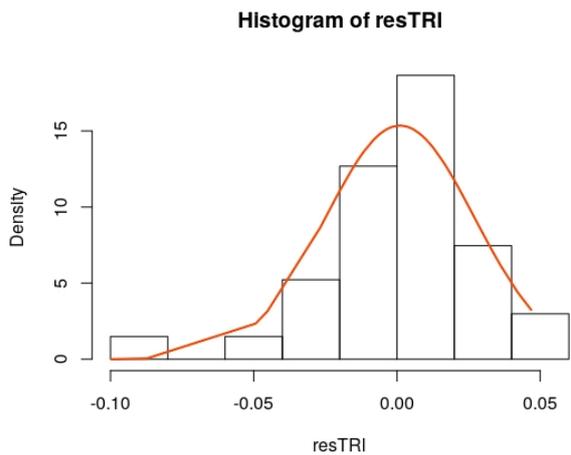
Possiamo notare come la banda blu sia più stretta rispetto alla verde, ma questo non ci sorprende poiché già sapevamo di non potere fidarci troppo del fatto che i residui fossero gaussiani.

1.7 Serie troncata

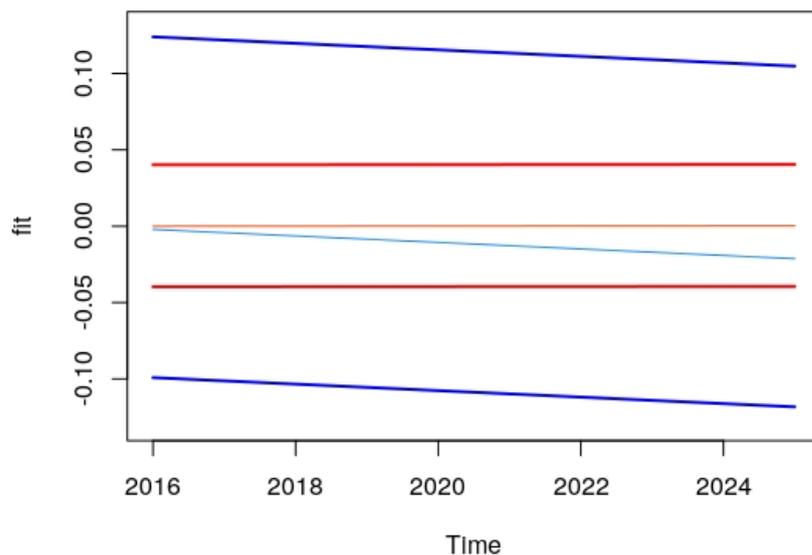
La previsione e i residui ottenuti nella sezione precedente potrebbero essere stati influenzati da un passato remoto in cui la serie assumeva comportamenti molto diversi rispetto al presente, che ha influenzato l'analisi. Per superare questa difficoltà abbiamo provato a considerare la serie solo a partire dal secondo dopoguerra e abbiamo proceduto con la previsione come per l'intera serie ottenendo il seguente grafico:



In analogia con quanto fatto in precedenza abbiamo in blu la banda data dai quantili empirici, in verde una banda ottenuta da quantili gaussiani, come vediamo queste sono praticamente sovrapposte (anche se, dando un'occhiata ai valori, si nota come la banda empirica risulti ancora una volta più stretta). Ciò accade perché i residui di questa serie sono molto migliori, nelle figure successive possiamo apprezzarne ACF e istogramma:



Tornando alla previsione osserviamo che questa sembra a prima vista costante, in realtà controllando i valori ottenuti si scopre che questa è leggermente in crescita contrastando la previsione ottenuta considerando tutta la serie. Un fatto positivo è che le bande siano decisamente più strette rispetto a prima, il pericolo è che potremmo averle ristrette troppo.

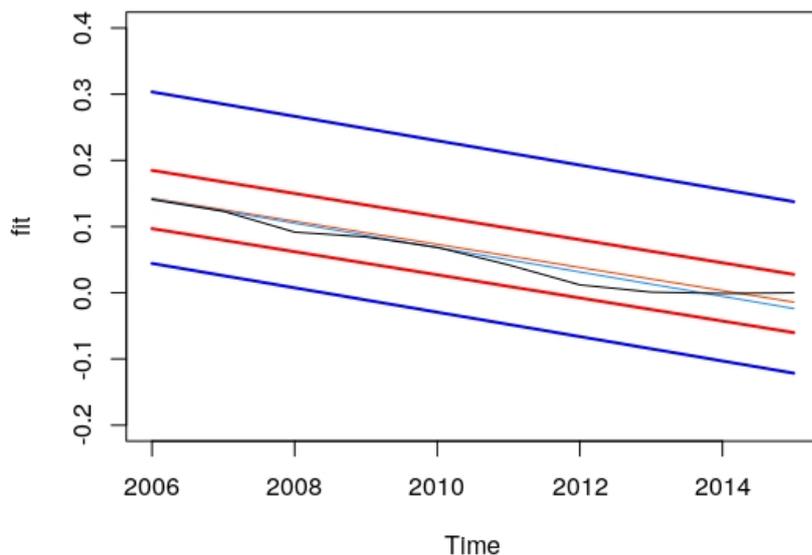


In effetti, come si vede nel grafico qui sopra, la banda empirica ottenuta dalla serie troncata, in rosso, contiene anche la previsione ottenuta tramite tutta la serie, in azzurro, ed è molto più stretta della banda

ottenuta con tutta la serie, in blu.

1.8 Autovalidazione

Un'ulteriore verifica della bontà delle bande calcolate tramite la serie troncata è stata fatta considerando la serie storica dal dopoguerra al 2005 ignorando in partenza i dati degli ultimi dieci anni al fine di vedere se questi fossero compatibili con le bande. Si è ottenuto il seguente grafico:



Osserviamo come quest'ultima figura sia molto simile a quella precedente ma si riferisce alla finestra temporale 2006 – 2015 mentre l'altra era riferita al futuro (2016 – 2025). I colori sono coerenti con la figura precedente nel senso che linee dello stesso colore sono state ottenute con lo stesso metodo, in più nell'ultima figura appare una linea nera che rappresenta l'andamento effettivo della serie negli ultimi 10 anni. Come si può apprezzare entrambe le previsioni si sono rivelate buone inoltre le bande di confidenza hanno funzionato più che bene. Dal punto di vista della previsione per gli anni 2016 – 2025 si può affermare che se la serie manterrà l'andamento regolare degli ultimi decenni (e abbiamo visto che in passato per smuovere la regolarità della serie ci sono voluti elementi di portata enormemente destabilizzante come le guerre mondiali), allora le bande di confidenza ottenute con la serie troncata e le previsioni (sia quella proveniente dalla serie troncata che quella proveniente dalla serie completa) sono affidabili. Un'ultimo commento può essere fatto riguardo il "pessimismo" delle nostre previsioni: come si può notare nell'ultimo grafico infatti, le analisi effettuate usando i dati fino al 2005 per prevedere l'ultimo decennio si sono rivelate tutte decrescenti, tuttavia l'analisi delle bande di confidenza non permette di escludere che nei prossimi anni possa avvenire un cambio di trend o almeno una crescita sporadica ed in effetti la previsione per i prossimi 10 anni effettuata tramite la serie troncata si è rivelata leggermente crescente. Quello che si può escludere con una certa sicurezza, per concludere, è una variazione sostanziale (sia in positivo che in negativo) del valore della moneta. Un fatto che probabilmente ha contribuito in maniera decisiva a questo appiattimento del trend è l'entrata in vigore dell'Euro, una moneta unica europea più forte si può supporre che sia in di natura più stabile.

Capitolo 2

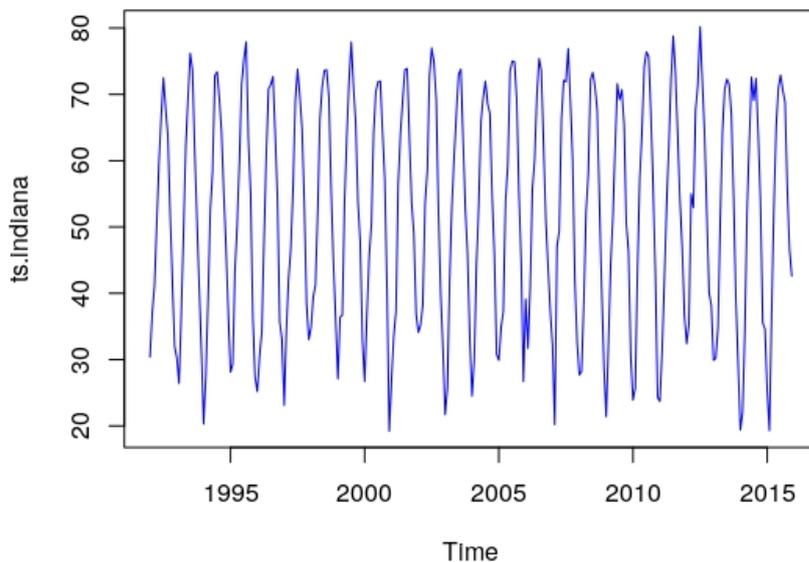
Temperature nello stato dell'Indiana

La seguente serie storica raccoglie i dati delle temperature medie mensili nello stato dell'Indiana, negli USA, dal gennaio del 1992 al dicembre del 2015. Sebbene fossero disponibili anche i dati relativi alla prima parte del 2016, si è preferito non adoperarli in quanto incompleti.

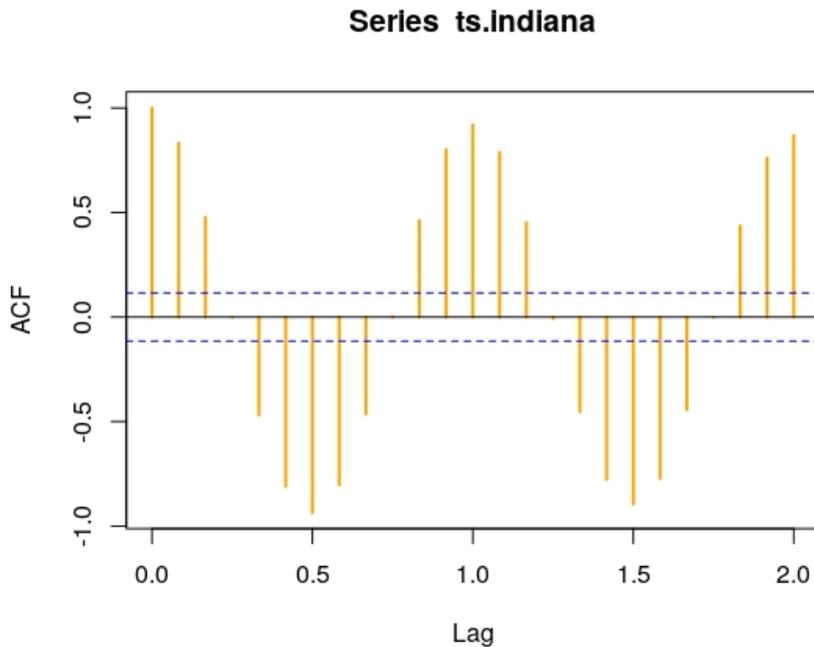
2.1 Grafico e autocorrelazione

Per la natura del fenomeno che vogliamo registrare ci aspettiamo una forte periodicità annuale e un trend molto prossimo ad essere costante, in quanto le temperature seguono ovviamente il ritmo naturale delle stagioni e non esistono validi motivi per cui si possa sperare in un aumento o una diminuzione delle temperature nel corso di così pochi anni.

Il grafico della serie storica risulta:



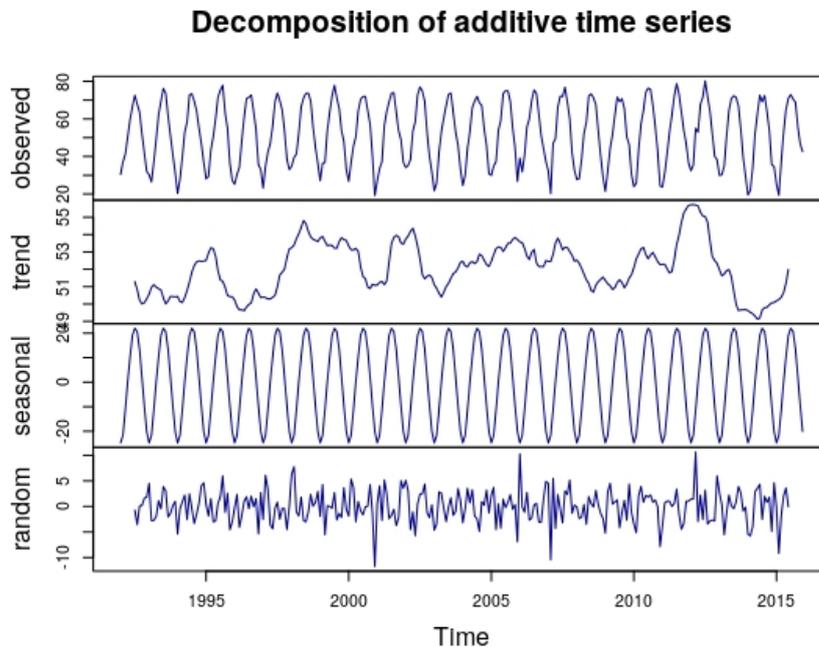
L'osservazione grafica della serie si allinea all'idea di base che avevamo, andiamo ora a studiare la funzione di autocorrelazione:



Non è difficile notare dei picchi positivi in corrispondenza del periodo e dei picchi negativi a metà di esso, entrambi segnali che confermano la natura fortemente stagionale della serie.

2.2 Decomposizione della serie

Poichè le oscillazioni mantengono nel tempo un'ampiezza piuttosto costante, sembra ragionevole preferire un modello additivo ad uno moltiplicativo per decomporre la serie. Il grafico che otteniamo è:

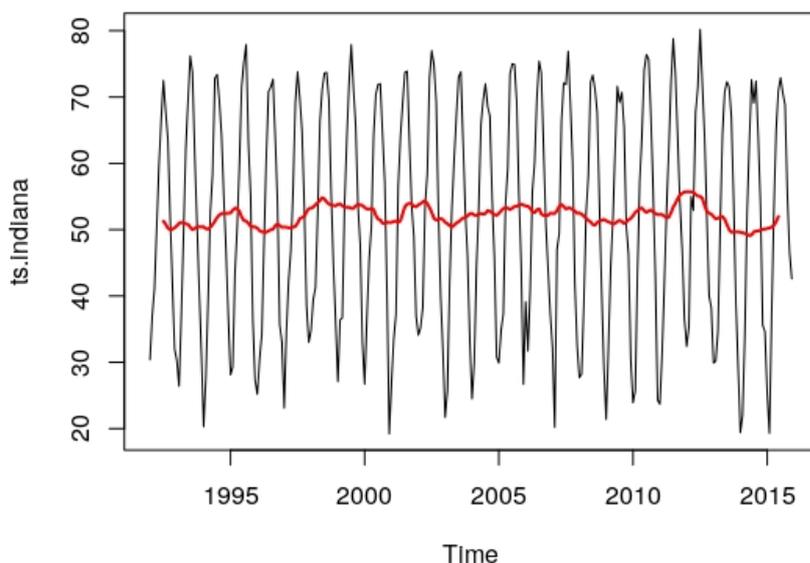


Come possiamo osservare, la decomposizione ci da risultati confortanti:

- il trend è stazionario in quanto le sue variazioni sono minime;
- la stagionalità è accentuata e dominante rispetto al trend;
- il residuo è moderato ma non del tutto trascurabile.

2.3 Analisi e previsione del trend

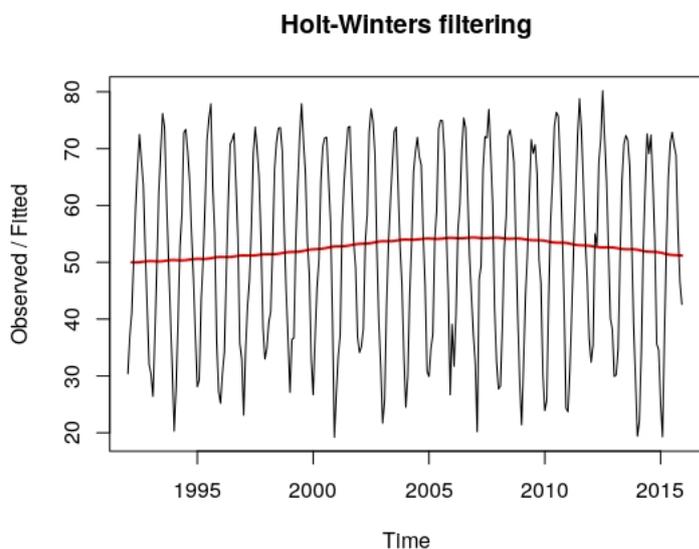
La decomposizione della serie vista nella sezione precedente era in linea con l'idea intuitiva che la serie fosse a dominanza stagionale; tuttavia, sebbene moderato, il trend ottenuto appare frastagliato, si può supporre che ciò accada per via di una certa quantità di rumore che viene erroneamente catturata in esso. A riprova di questo fatto può essere utile disegnare il trend ottenuto dal comando *decompose* sovrapposto a quello della serie storica:



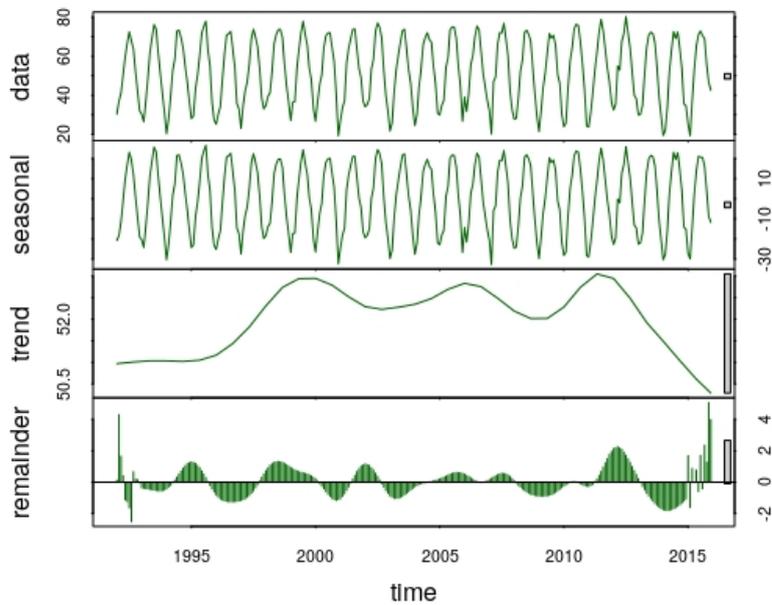
Analisi del trend con Holt-Winters Poichè il nostro intento è di catturare il trend tralasciando momentaneamente la stagionalità, ci siamo affidati al metodo *SET* piuttosto che ad Holt-Winters completo, inoltre i parametri e le condizioni iniziali sono stati modificati rispetto a quelli standard e, dopo numerosi tentativi si è posto:

- $\alpha = 0.001$ in quanto non vogliamo che il risultato aderisca troppo alla serie;
- $l.start = 50$ che è vicino al valore medio dei dati;
- $b.start = 0$.

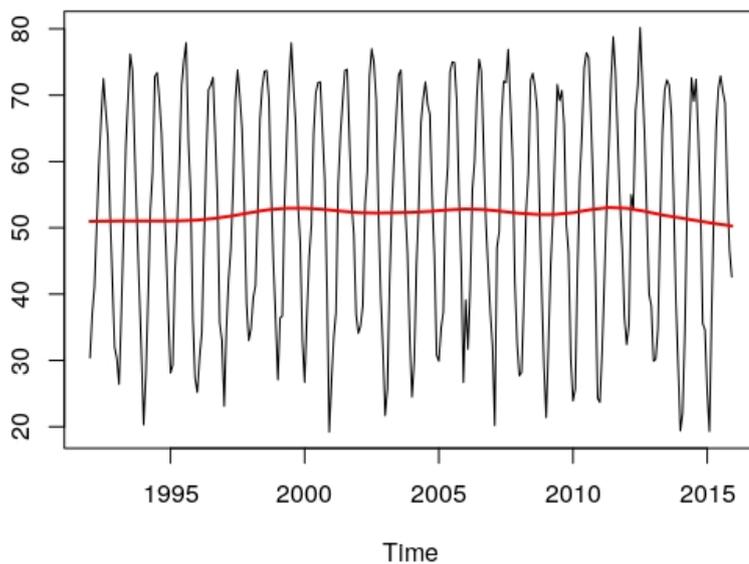
Il grafico ottenuto è il seguente:



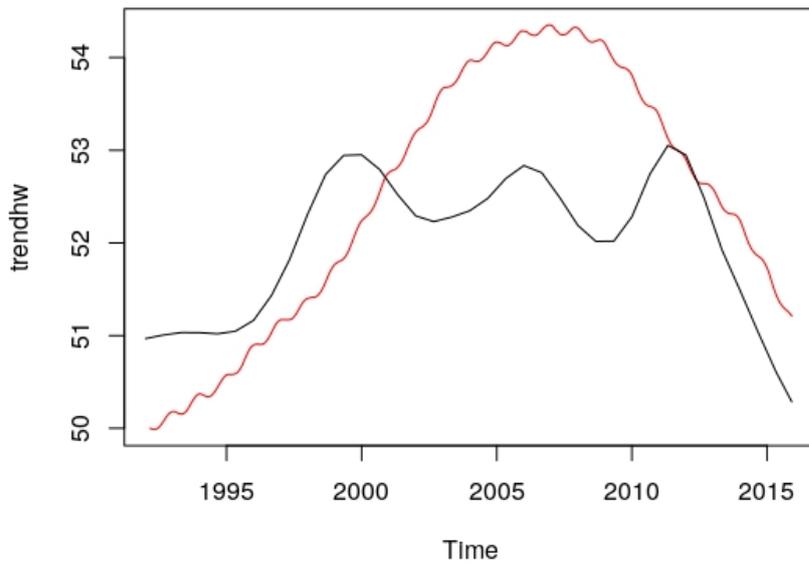
Analisi del trend col metodo STL Un'altra strada fruttuosa per catturare il trend è quella di utilizzare la funzione *stl* con un parametro *k* basso, in particolare il risultato che ci soddisfa di più è stato ottenuto con $k = 2$, di cui riportiamo il grafico qui di seguito



Se adesso, come abbiamo fatto per *decompose*, disegniamo il grafico del trend ottenuto sopra quello della serie, possiamo apprezzare come questo risulti molto più regolare e perciò più vicino alla nostra idea intuitiva di trend.

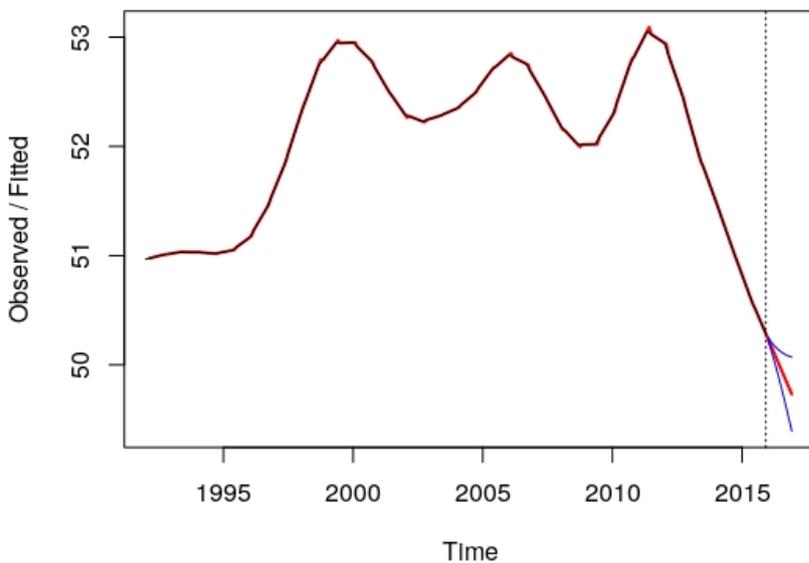


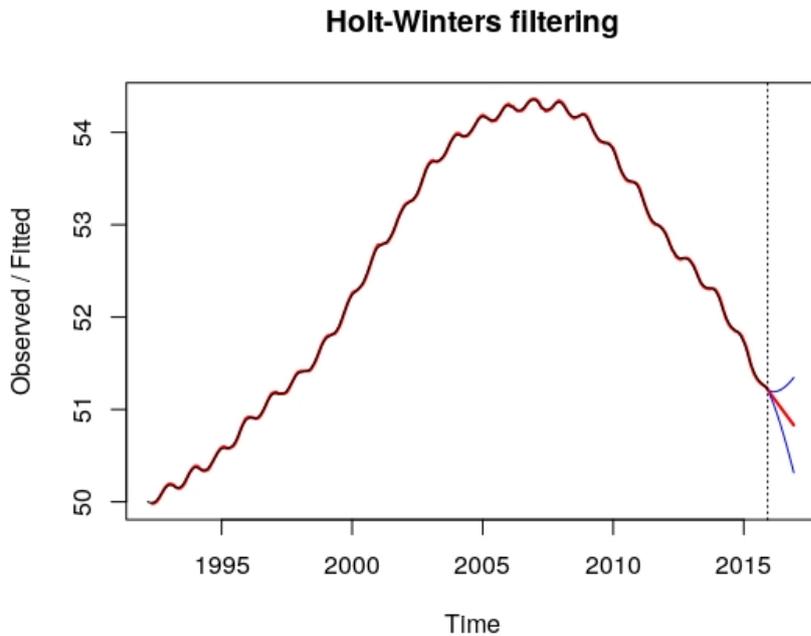
Confronto fra i trend e previsione Sovrapponendo i grafici dei due trend si ottiene la figura:



Non si può dire, ad occhio, quale dei due sia il migliore. Quello ottenuto col SET, in rosso, presenta delle leggere ondulazioni causate dalla forte stagionalità delle serie, ma a parte queste ondulazioni si può registrare un solo cambio di trend. Quello ottenuto con la funzione *stl*, in nero, non presenta ondulazioni ma di contro subisce diversi cambi di trend. Un difetto comune ad entrambi è questa discesa finale, probabilmente dovuta al fatto (intuitivamente casuale) che negli ultimi tre anni i picchi, sia quelli positivi sia quelli negativi, sono stati più bassi rispetto agli anni precedenti. Applicando ad entrambi i trend il metodo SET ed effettuando una previsione per i successivi 12 mesi si ottengono i due seguenti grafici:

Holt-Winters filtering

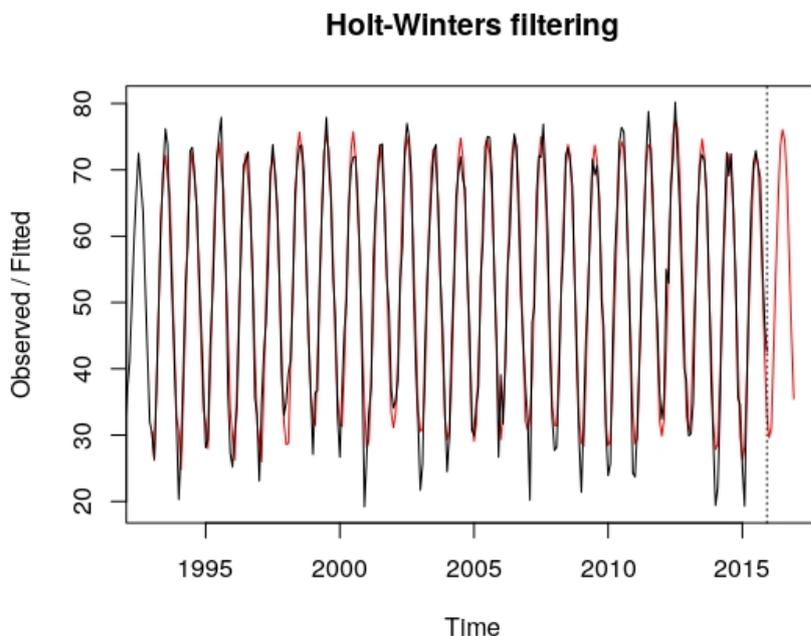




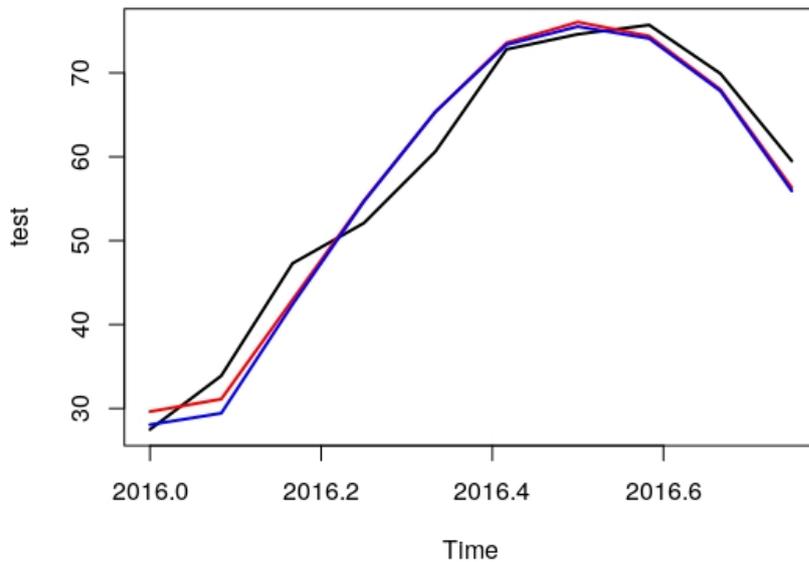
Proprio per il fatto che entrambi i trend subiscono una discesa finale, le due previsioni risultano non troppo esaltanti in quanto è ragionevole supporre che la decrescita negli ultimi tre anni sia piuttosto casuale e perciò sembra lecito attendersi un ulteriore cambio di trend. In quest'ottica un risultato migliore è rappresentato, in entrambe le previsioni, dalle bande di confidenza superiori.

2.4 Previsione della serie

Con lo scopo di prevedere l'andamento della serie per l'anno 2016 ci siamo affidati nuovamente ad Holt-Winters, senza modificare i parametri iniziali otteniamo già un buon risultato:

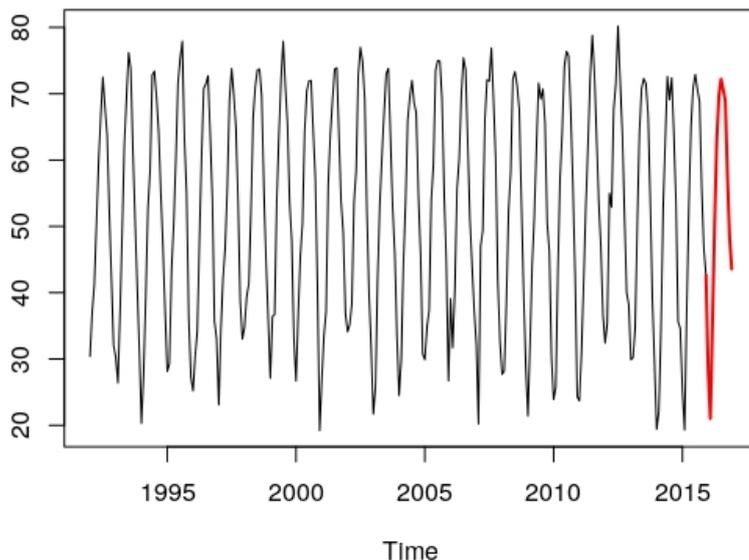


Come vediamo, la previsione, in rosso, in qualche modo "media" i picchi della serie, nel senso che si trova poco sopra i picchi più bassi e poco sotto quelli più alti, mentre nei tratti intermedi è molto aderente alla serie. Per verificare la bontà della previsione abbiamo confrontato quest'ultima con i dati relativi ai primi 10 mesi del 2016 che inizialmente avevamo messo da parte, il risultato è il seguente:

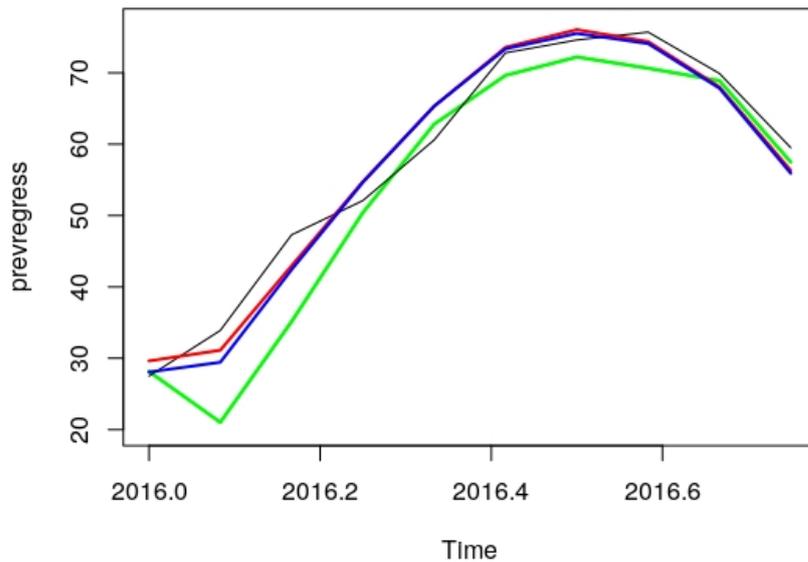


In nero possiamo vedere il disegno della serie nel 2016, in blu la previsione di Holt-Winters con i parametri standard e in rosso una previsione effettuata sempre con Holt-Winters ma con parametri (leggermente) modificati a posteriori, nel senso che abbiamo adattato i parametri in modo che la previsione fosse più aderente (almeno ad occhio) ai dati "futuri" disponibili; non è possibile sapere, ovviamente, se i parametri ottenuti in questo modo potranno predire i dati non ancora disponibili meglio dei parametri standard.

Buoni risultati in termini di previsione sono stati raggiunti anche tramite metodi autoregressivi. In parte effettuando l'eliminazione di alcuni fattori precedentemente considerati, in parte basandoci sulla pura intuizione siamo giunti ad un modello soddisfacente considerando come fattori di regressione quelli ottenuti traslando la serie di rispettivamente di un mese e di un anno; i valori R^2 e Adjusted R^2 ottenuti in questo modo sono entrambi attorno al 92%, il che è confortante. In effetti la previsione ottenuta grazie al modello autoregressivo (mostrata nel grafico successivo) appare ad occhio abbastanza ragionevole.



Andiamo ora a confrontare questa previsione con i dati del 2016 e con le previsioni di Holt-Winters:



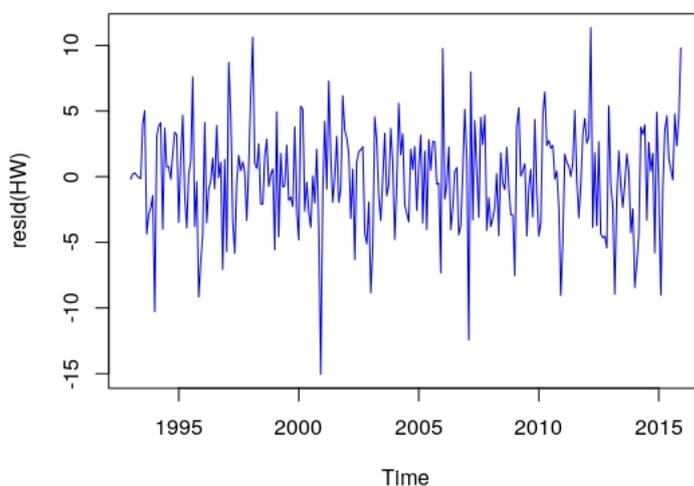
Quello che si può vedere è che il modello autoregressivo è leggermente peggiore di Holt-Winters per i primi dati, ma non del tutto inappropriato, inoltre il fatto che il secondo dato sia stato sottostimato da tutte e tre le previsioni, potrebbe essere dovuto ad un febbraio 2016 inaspettatamente caldo, più che ad un'inappropriatezza dei modelli.

2.5 Analisi dei residui

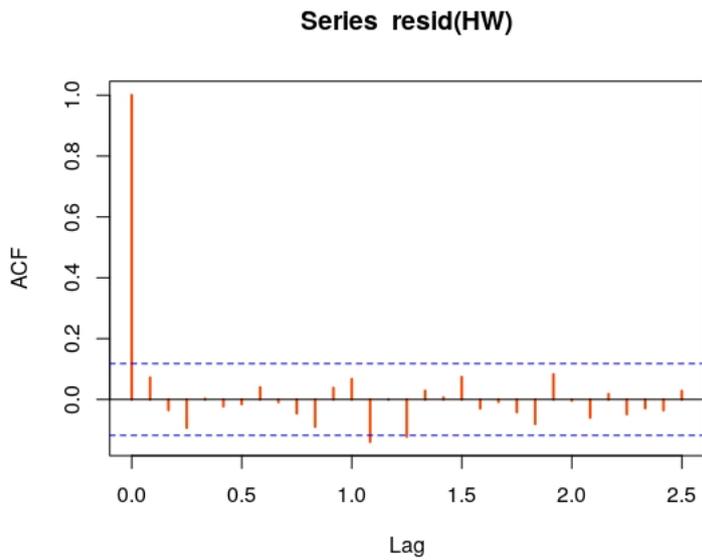
Ricapitolando, per lo studio della serie sono stati utilizzati i seguenti metodi:

- decomposizione additiva
- Holt-Winters
- autoregressione.

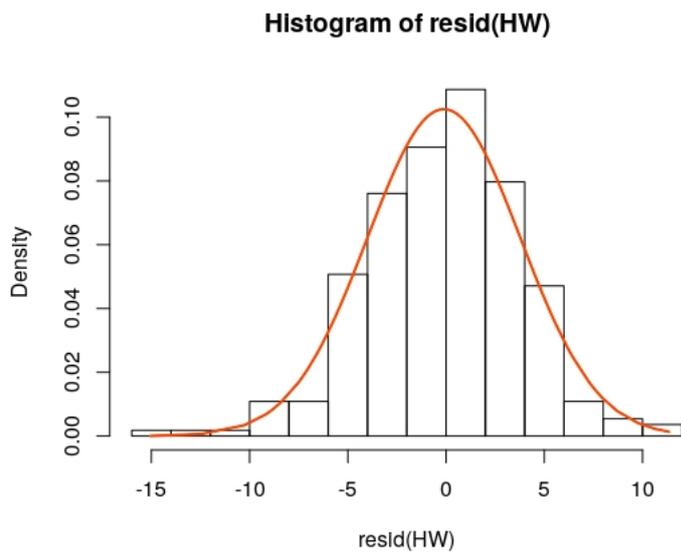
Tutti e tre i metodi hanno residui contenuti e varianza spiegata oltre il 90%, questo non ci sorprende in quanto già ad occhio si intuisce che la serie non presenta un rumore eccessivo. Tuttavia osservando le *acf* dei residui dei tre metodi il migliore è risultato quello di Holt-Winters con i parametri leggermente modificati su cui ci concentreremo in seguito e di cui qui riportiamo il grafico dei residui



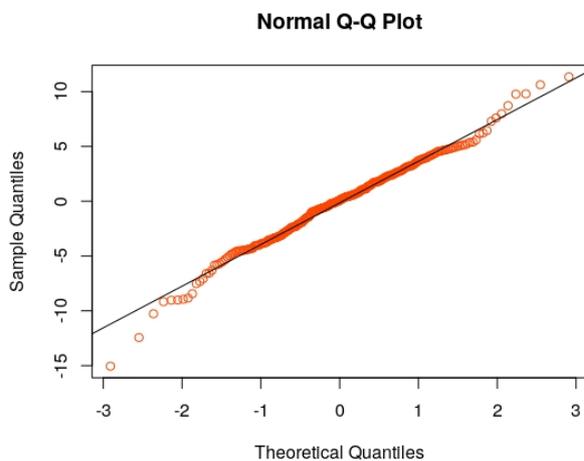
e la funzione di autocorrelazione



Un altro aspetto interessante dei residui si evince dall'istogramma:

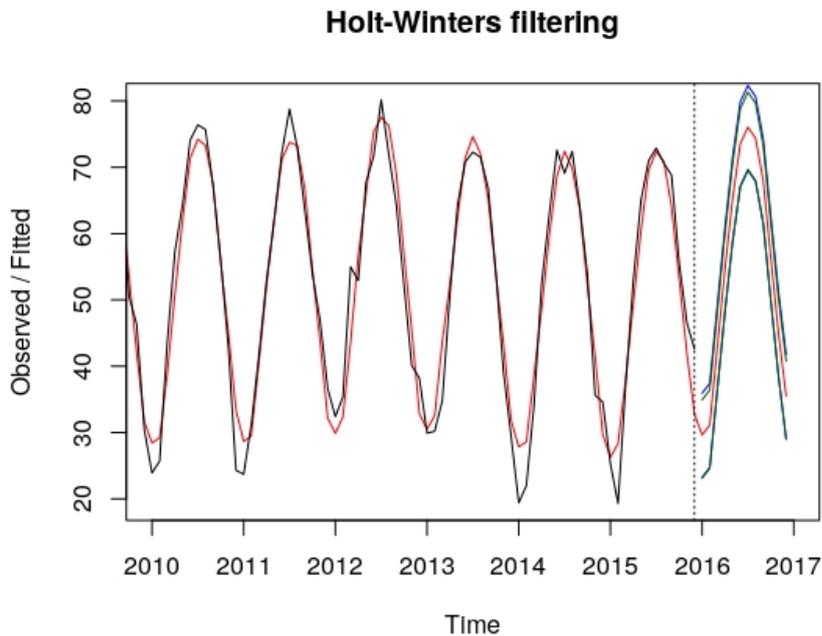


Disegnando sopra di esso la densità gaussiana definita dalla media e dalla deviazione standard dei dati, possiamo apprezzare come l'istogramma segua grosso modo la campana, il che ci fa supporre che effettivamente i residui abbiano natura gaussiana. Il *qqplot* conferma questa ipotesi almeno per valori compresi fra -1 e 1 :

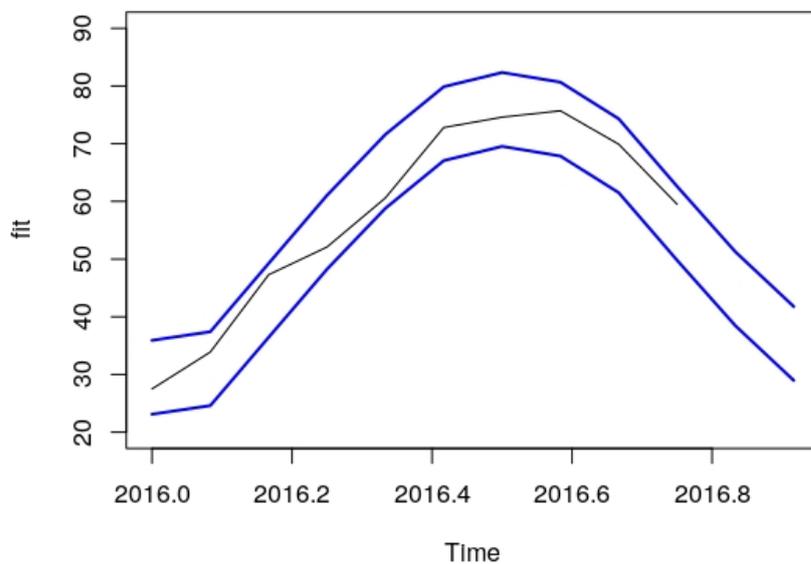


2.6 Bande di confidenza della previsione

I risultati ottenuti nella sezione precedente ci suggeriscono l'utilizzo di quantili parametrici gaussiani per la costruzione di una banda di confidenza per la previsione generata da Holt-Winters.



Nella figura precedente si può apprezzare infatti come la banda di confidenza teorica, in blu, e quella empirica, in verde, entrambe al 90% siano praticamente indistinguibili. Per concludere riportiamo il seguente grafico



che ci mostra come i dati del 2016 siano stati finora dentro la banda di confidenza generata dai quantili parametrici gaussiani.