**Modelli Statistici e Statistical Learning - 2022/2023**

RELAZIONE ESERCIZIO 1

Gruppo SparadAIs



L'esercizio consiste nello stimare il trend e la componente stagionale di due serie storiche. La prima serie storica è trimestrale e rappresenta il tasso di disoccupazione tra i maschi, mentre la seconda serie mensile riguarda l’indice di produzione industriale.

**Serie trimestrale: Tasso di disoccupazione tra i maschi**

Consideriamo i dati sul tasso di disoccupazione in Italia della popolazione maschile nel periodo che va dal quarto trimestre del 1992 al quarto trimestre del 2020.

Leggiamo i dati in R, assegnando ad ogni valore un periodo di osservazione. In particolare, specifichiamo il periodo di partenza della serie (4° trimestre del 1992) e la frequenza “trimestrale”.

Immagine che contiene testo, screenshot, arancia

Descrizione generata automaticamente



Iniziamo stimando il trend della serie storica. Il grafico suggerisce che la serie osservata, al variare del tempo, presenta una tendenza di fondo del tipo .

Utilizziamo il modello di regressione lineare per stimare un modello che interpreti la tendenza di fondo della serie. In particolare, usiamo un modello in cui la componente deterministica è un polinomio di ordine e la componente stocastica è generata da un processo white noise. Scegliamo l’ordine del polinomio utilizzando il criterio basato sull’indice di determinazione corretto .

Immagine che contiene testo

Descrizione generata automaticamente

Immagine che contiene testo

Descrizione generata automaticamente

Iniziamo stimando un Trend polinomiale di ordine 1, cioè .

Consideriamo α = 0.05. Il test di Fisher presenta un p-value < α, di conseguenza rifiutiamo l’ipotesi nulla , cioè possiamo affermare che i regressori complessivamente considerati sono statisticamente significativi nello spiegare . In particolare, dai test marginali di entrambi i coefficienti di regressione e notiamo che il p-value << α, quindi sia l’intercetta che il regressore t sono entrambi fortemente significativi nello spiegare .

L’ultima considerazione riguarda l’ che presenta un valore molto basso, .

Immagine che contiene testo

Descrizione generata automaticamente

Dal confronto tra i valori osservati e i valori stimati (con =1) notiamo che il modello non si adatta ai dati. Seguendo il criterio basato sull'indice di determinazione corretto proseguiamo nella stima del modello con ordine superiore.

Immagine che contiene testo

Descrizione generata automaticamente

Immagine che contiene testo

Descrizione generata automaticamente

Anche in questo caso, il test di Fisher presenta un p-value < α, di conseguenza almeno un regressore è statisticamente significativo nello spiegare . I test marginali dei coefficienti di regressione , presentano un p-value << α, quindi tutti i regressori sono fortemente significativi nello spiegare .

L’ aumenta rispetto al caso precedente, .

Immagine che contiene testo

Descrizione generata automaticamente



Confrontando i valori osservati e i valori stimati (con =2) osserviamo che anche in questo caso il modello non si adatta molto bene ai dati, ma l’andamento quadratico è sicuramente migliore rispetto al caso precedente nello spiegare i dati.

Continuiamo nella stima del modello con ordine 3.

Immagine che contiene testo

Descrizione generata automaticamente

Immagine che contiene testo

Descrizione generata automaticamente

Il p-value del test di Fisher è minore di α, anche in questo caso rifiutiamo l’ipotesi nulla . I test marginali suggeriscono che tutti i regressori sono fortemente significativi nello spiegare , essendo p-value < α in tutti e quattro i casi.

L’ aumenta lievemente rispetto al caso precedente, .

Immagine che contiene testo

Descrizione generata automaticamente



L’adattamento ai dati migliora ancora.

Proseguiamo stimando un trend polinomiale di ordine 4.

Immagine che contiene testo

Descrizione generata automaticamente

Immagine che contiene tavolo

Descrizione generata automaticamente

Essendo il p-value del test di Fisher minore di α, possiamo affermare che tutti i regressori complessivamente considerati sono significativi nello spiegare la variabile dipendente. Anche in questo caso in tutti i test marginali rifiutiamo l’ipotesi secondo cui con i=0,1,2,3,4.

L’ aumenta notevolmente rispetto al caso precedente, .

Immagine che contiene testo

Descrizione generata automaticamente

Proseguiamo con la stima del modello con ordine superiore per capire se possiamo ottenere un corretto maggiore rispetto al caso corrente.

Immagine che contiene testo

Descrizione generata automaticamente

Immagine che contiene tavolo

Descrizione generata automaticamente

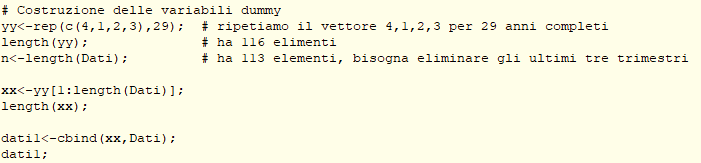
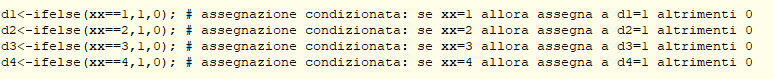
Osserviamo che, con un polinomio di grado =5, il coefficiente di regressione non è significativo nello spiegare .

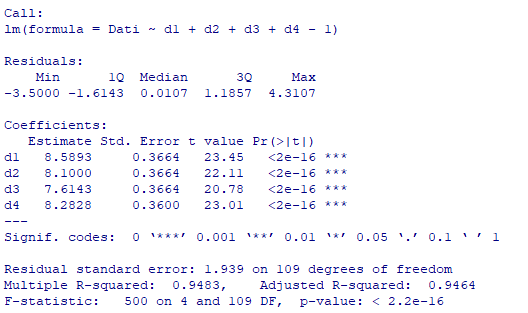
Inoltre, L’ diminuisce rispetto al caso precedente, .

Quindi, secondo il criterio basato sull’indice di determinazione corretto scegliamo la specificazione con un polinomio di ordine 4. Il modello stimato è quindi:

Per quanto riguarda il trend, concludiamo che il modello polinomiale di ordine 4 si adatta ragionevolmente bene ai dati osservati. Tuttavia, bisogna notare che il modello stimato non riesce a catturare perfettamente i picchi della serie. Attraverso la stima della stagionalità proviamo a superare questo limite.

Quindi proseguiamo stimando la stagionalità dalla serie storica utilizzando le variabili dummy. Nel nostro caso, essendo la serie trimestrale, abbiamo necessità di utilizzare quattro variabili dummy, ognuna delle quali indica il trimestre dell’anno in cui sono stati osservati i dati.





Esattamente come ci aspettavamo, tutte le variabili dummy risultano essere significative nel modello che stima la stagionalità. L’ assume un valore elevato.

Di seguito è riportato un confronto tra i dati osservati e la stima della stagionalità:

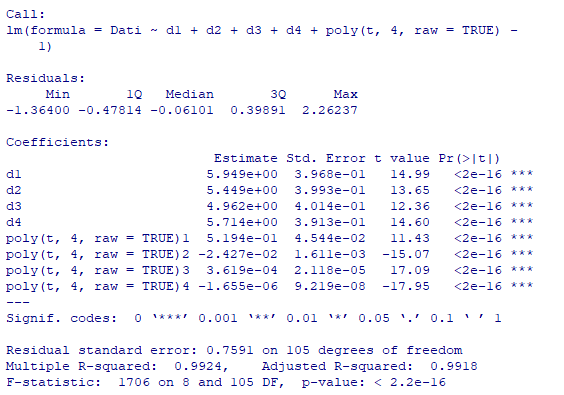




L'andamento della componente stagionale studia bene la periodicità. A questo punto, consideriamo contemporaneamente componente stagionale e il trend per ottenere il modello complessivo che descrive sia la periodicità che la tendenza della serie storica.

Il modello finale che considera sia la componente stagionale che il trend è il seguente:





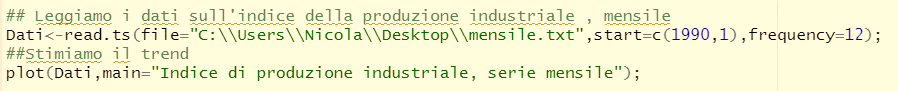
Si può notare che sia le variabili dummy sia la componente polinomiale sono fortemente significativi. Otteniamo inoltre un valore dell’ più alto rispetto a quelli ottenuti precedentemente stimando singolarmente stagionalità e trend. Concludiamo che stima stagionale e trend combinati consentono di avere una migliore spiegazione di variabilità del modello rispetto ai casi singoli.

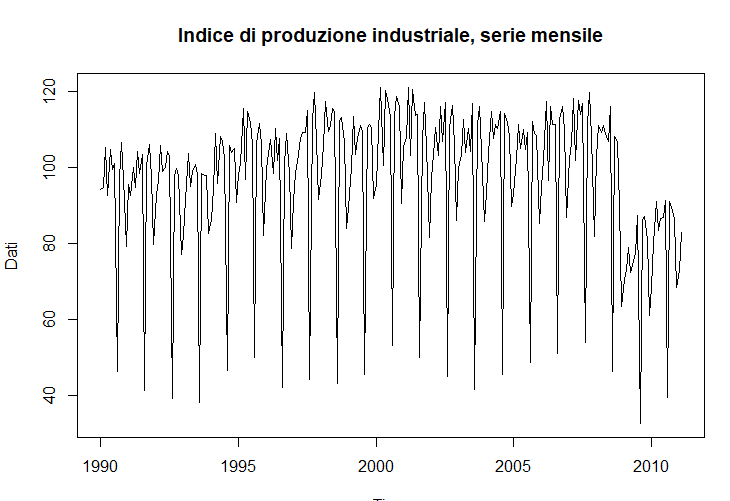




**Serie mensile: Indice della produzione industriale**

Consideriamo i dati sull’indice di produzione industriale nel periodo gennaio 1990 – febbraio 2011.

Come già fatto precedentemente, leggiamo i dati in R specificando il periodo di partenza della serie (1° mese del 1990) e la frequenza “mensile” ovvero 12 misure in un anno.



Anche in questo caso, iniziamo con la stima del trend della serie storica. Contrariamente a quanto visto con la serie trimestrale, il grafico non suggerisce una tendenza di fondo in particolare, ma sembra essere periodica.

Costruiamo la sequenza 1,2,…,T dove T è la lunghezza della serie:



Iniziamo stimando un Trend polinomiale di ordine 1, cioè :



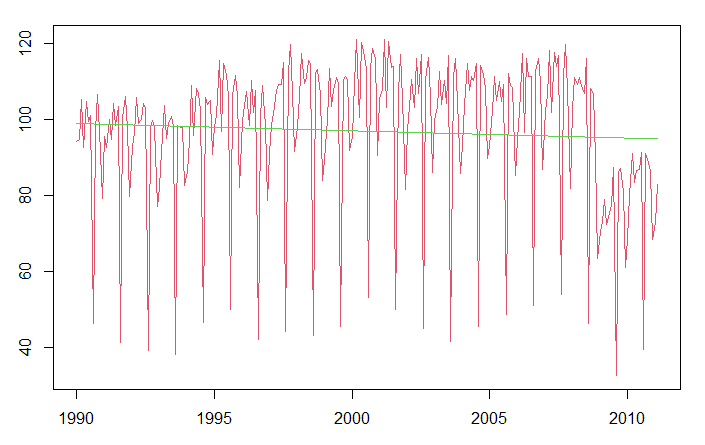
Immagine che contiene testo, ricevuta

Descrizione generata automaticamente

Il Test di Fisher coincide con quello marginale, β1 non è significativo.

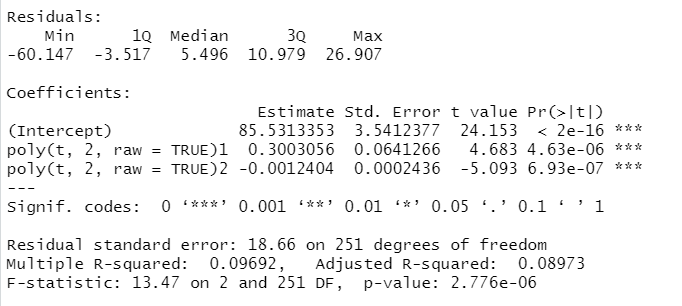
Confrontiamo il modello formulato con i dati:





Dal confronto tra i valori osservati e i valori stimati notiamo che il modello non si adatta ai dati. Ancora una volta, utilizziamo il criterio basato sull'indice di determinazione corretto , quindi proseguiamo nella stima del modello con ordine superiore.





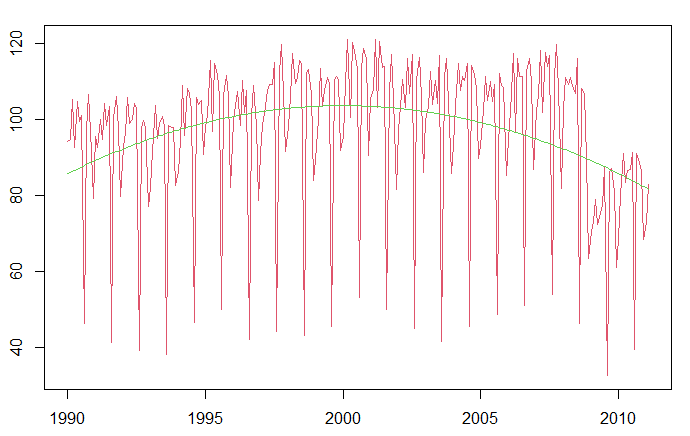
Questa volta il test di Fisher presenta un p-value < α, di conseguenza almeno un regressore è statisticamente significativo nello spiegare . In particolare, come si osserva dai test marginali dei coefficienti di regressione , , i p-value << α, quindi tutti i regressori sono fortemente significativi nello spiegare .

Otteniamo un migliore rispetto al caso precedente, . Tuttavia, osserviamo che si tratta di un valore poco soddisfacente.

Confrontiamo il modello stimato e i dati:

Immagine che contiene testo

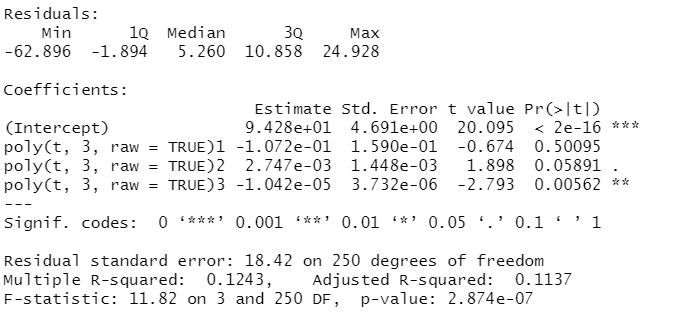
Descrizione generata automaticamente



Anche in questo caso il modello non si adatta molto bene ai dati, ma l’andamento quadratico è sicuramente migliore rispetto al caso precedente. Inoltre, il valore β2<0 rende la curva rivolta verso il basso.

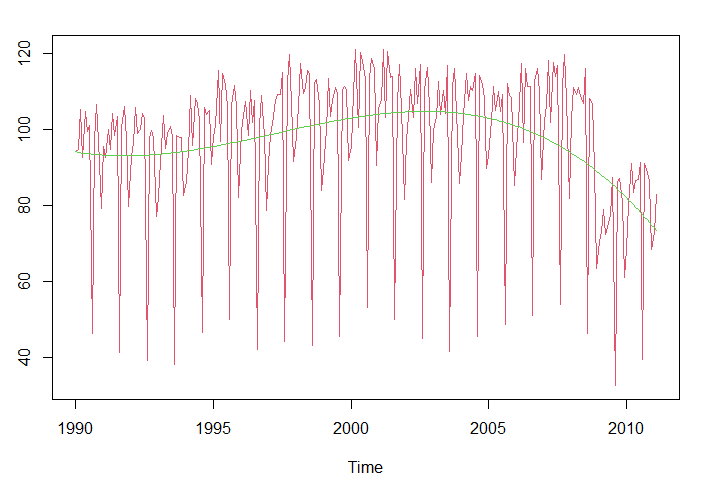
Continuiamo nella stima del modello con ordine 3.





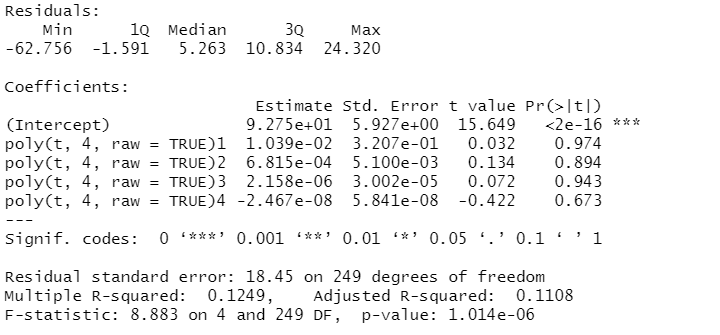
Il p-value del test di Fisher è minore di α, anche in questo caso rifiutiamo l’ipotesi nulla . In questo caso osserviamo anche che risulta essere per niente significativo mentre è leggermente significativo.

L’ aumenta lievemente rispetto al caso precedente, .

Immagine che contiene testo

Descrizione generata automaticamente

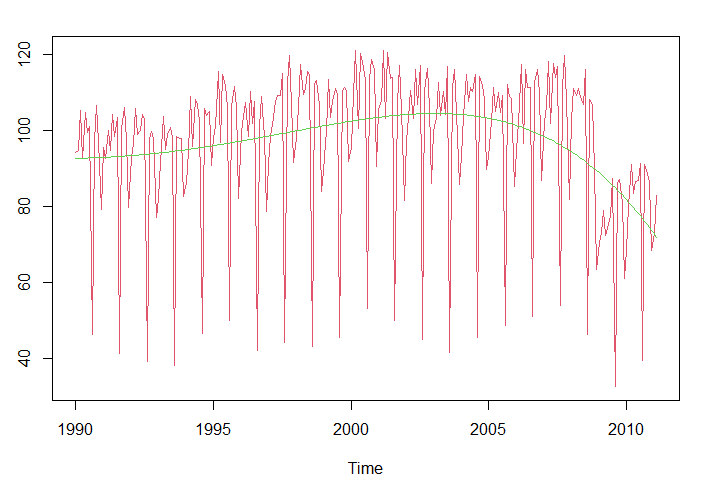
Il modello si adatta leggermente meglio, ma il criterio indica che si può proseguire stimando un modello di ordine 4.



L’ diminuisce rispetto al caso precedente, secondo il criterio bisogna fermarsi e non progredire ulteriormente.

Immagine che contiene testo

Descrizione generata automaticamente



Quindi, secondo il criterio basato sull’indice di determinazione corretto  scegliamo la specificazione con un polinomio di ordine 3. Il modello stimato è quindi:

Rispetto alla serie trimestrale, nella quale il solo trend si adattava discretamente bene ai dati osservati, nel caso della serie mensile, il modello polinomiale non si adatta in modo soddisfacente ai dati. In questo caso, la stima della stagionalità assumerà un’importanza ancora più rilevante.

Proseguiamo stimando la stagionalità dalla serie storica utilizzando le variabili dummy. Siccome questa volta la serie è mensile, abbiamo necessità di utilizzare 12 variabili dummy, ognuna delle quali indica il mese dell’anno in cui sono stati osservati i dati.

Immagine che contiene tavolo

Descrizione generata automaticamenteCostruiamo la stima ai minimi quadrati della componente stagionale:



Risulta opportuno inserire -1 nella specifica del modello per eliminare l’intercetta ed evitare la trappola delle dummy, scampando così la compromissione dell’adattabilità del modello.

Immagine che contiene tavolo

Descrizione generata automaticamente

Le variabili dummy risultano essere tutte molto significative; abbiamo inoltre un R2 corretto molto elevato. Confrontiamo il modello con i dati:

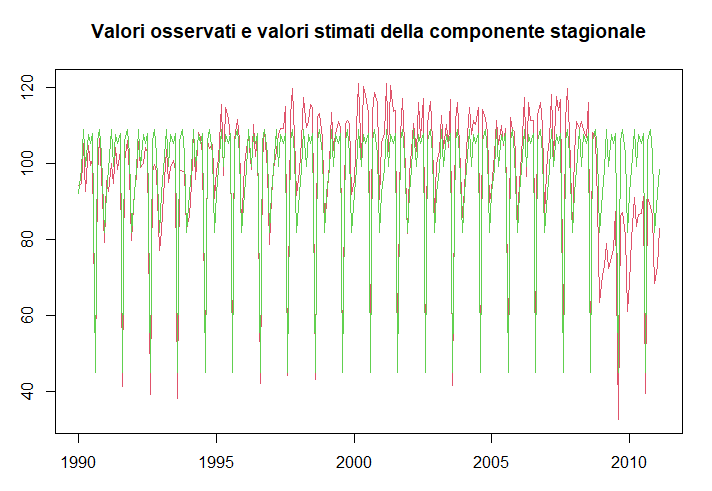
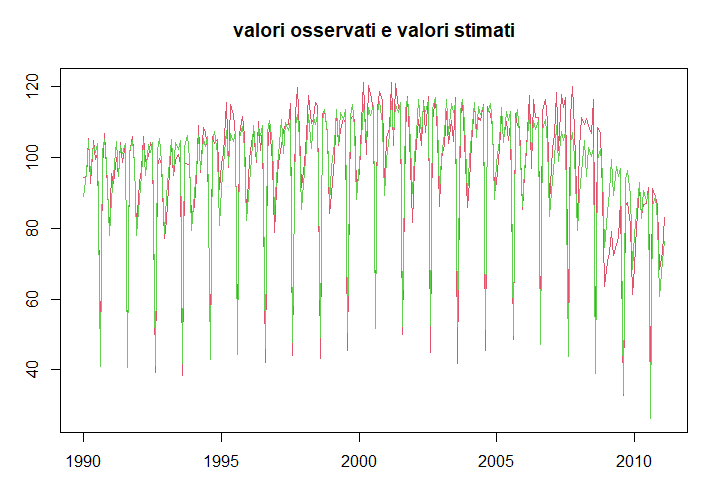


Immagine che contiene tavolo

Descrizione generata automaticamenteL'andamento della componente stagionale è più o meno conforme ai dati, studia bene la periodicità, ovviamente i nostri dati presentano anche la componente del trend. Stimiamo dunque il modello finale, che comprende sia la componente trend che la componente stagionale:

Otteniamo un molto alto pari a 0.9965 e anche superiore rispetto all’ della sola componente stagionale, il che significa che il modello spiega quasi tutta la variabilità osservata. Confrontiamo a questo punto i dati con il modello finale: