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

RELAZIONE ESERCIZIO 1

Gruppo SparadAIs



// aggiustare questa parte iniziale

L'esercizio consiste nello stimare il trend e la componente stagionale di due serie storiche. La prima serie storica è trimestrale e “rappresenta” i dati sul 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 tra la popolazione maschile nel periodo che va dal quarto trimestre del 1992 al quarto trimestre del 2022.

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 generato da un processo di white noise. Scegliamo l’ordine r 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 .

Dall'analisi dei residui notiamo che la mediana, pari a 0.0912, risulta essere maggiore della media (che ipotizziamo nulla per le ipotesi fondamentali del modello lineare). Di conseguenza concludiamo che la curva normale dei residui è leggermente asimmetrica a destra. // controllare non so se è giusto

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 y. 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 y.

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 r=1) notiamo che il modello non si adatta molto bene 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 in seguito il test di Fisher presenta un p-value < α, di conseguenza almeno un regressore è statisticamente significativo nello spiegare y. I test marginali dei coefficienti di regressione , presentano un p-value << α, quindi tutti i regressori sono fortemente significativi nello spiegare y.

L’ aumenta rispetto al caso precedente, .

Immagine che contiene testo

Descrizione generata automaticamente



Confrontando i valori osservati e i valori stimati (con r=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 y, essendo p-value < α in tutti e quattro i casi .

L’ aumenta 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 ancora rispetto al caso precedente, .

Immagine che contiene testo

Descrizione generata automaticamente



Il modello polinomiale di ordine 4 si adatta bene ai dati osservati, anche se bisogna notare che non si riescono a studiare bene i picchi della serie.

Proseguiamo con la stima del modello con ordine superiore.

Immagine che contiene testo

Descrizione generata automaticamente

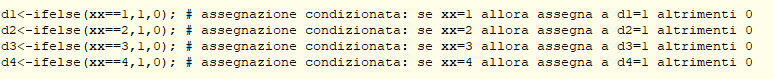
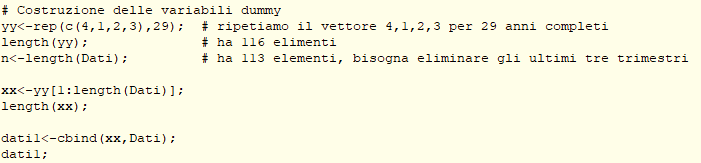
Immagine che contiene tavolo

Descrizione generata automaticamente

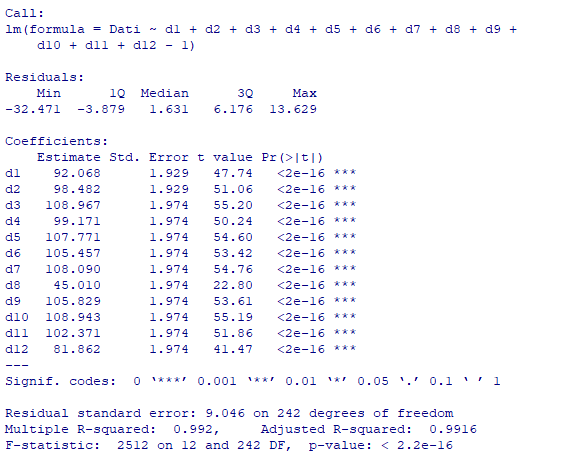
L’ diminuisce rispetto al caso precedente, .

Secondo il criterio basato sull’indice di determinazione corretto scegliamo la specificazione con un polinomio di ordine 4. Il modello stimato è 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, ogni delle quali indica il trimestre dell’anno in cui è stata osservata la serie.





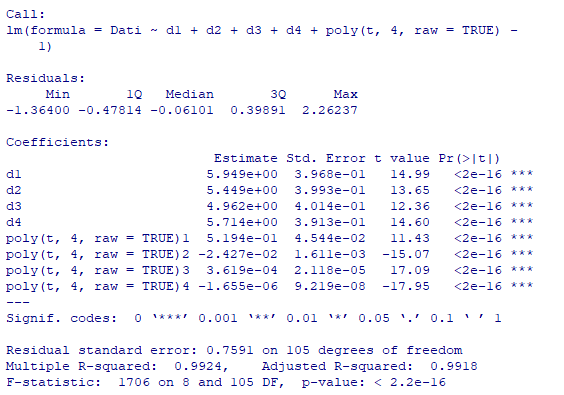






L'andamento della componente stagionale è più o meno conforme ai dati, studia bene la periodicità, ovviamente i nostri dati presentano anche la componente del trend.





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





Il modello si adatta più o meno bene ai dati, tuttavia non riusciamo a stimare bene i picchi.