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

RELAZIONE ESERCIZIO 2

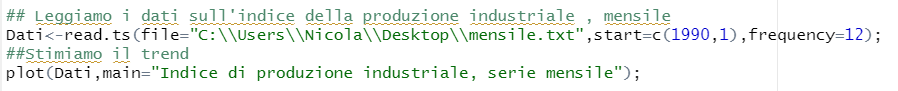
Gruppo SparadAIs

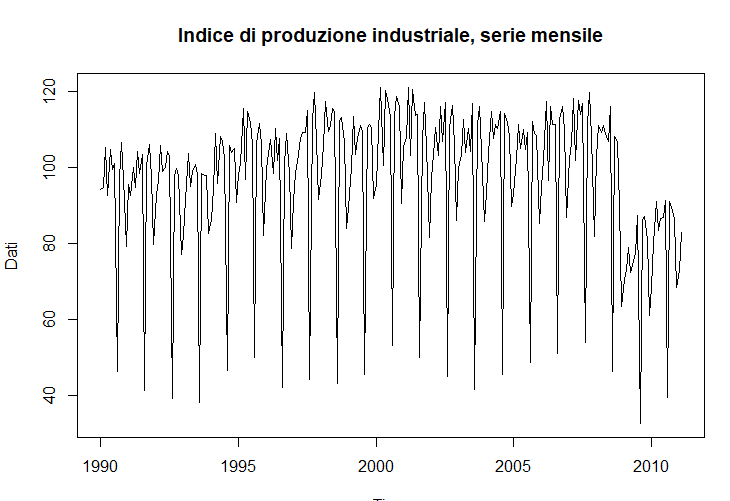


L'esercizio consiste nello stimare il trend e la componente stagionale di due serie storiche. La prima serie storica è trimestrale e rappresenta il tasso di disoccupazione tra i maschi, mentre la seconda serie mensile riguarda l’indice di produzione industriale.

**Serie mensile: Indice della produzione industriale**

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

Leggiamo i dati in R, assegnando ad ogni valore un periodo di osservazione. In particolare, specifichiamo il periodo di partenza della serie (1° mese del 1990) e la frequenza “mensile” ovvero 12 misure in un anno.



Iniziamo stimando il trend della serie storica. Il grafico non suggerisce alcuna tendenza di fondo, 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 :

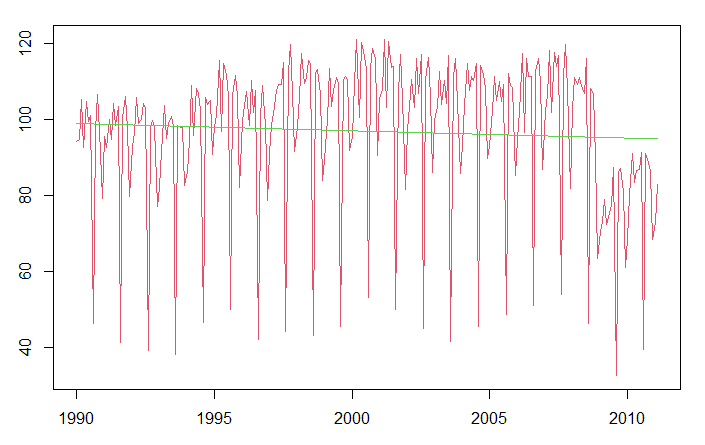


Immagine che contiene testo, ricevuta

Descrizione generata automaticamente

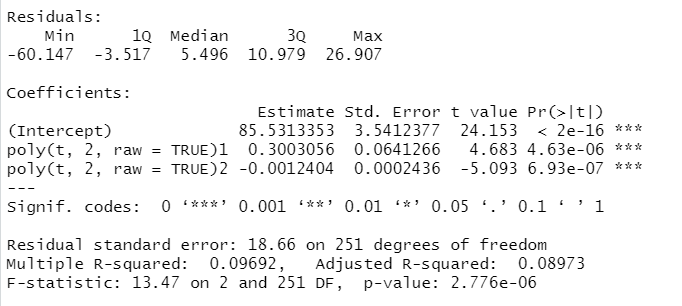
Dall’analisi dei residui notiamo che la curva normale dei residui è leggermente asimmetrica a sinistra, il Test di Fisher coincide con quello marginale, β1 non è significativo.Confrontiamo il modello formulato con i dati:





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



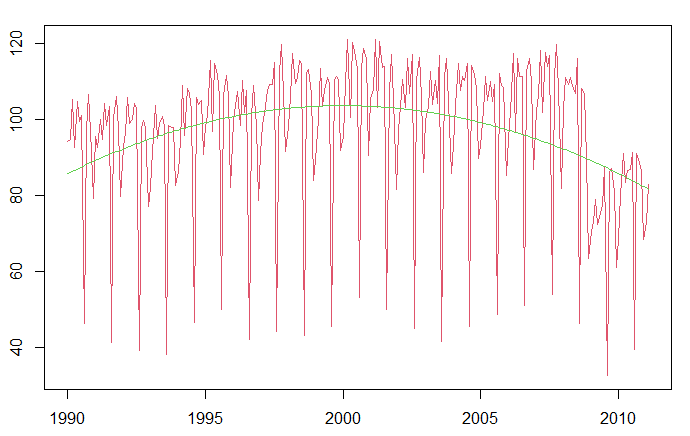


Anche in questo caso, il test di Fisher presenta un p-value < α, di conseguenza almeno un regressore è statisticamente significativo nello spiegare . I test marginali dei coefficienti di regressione , presentano un p-value << α, quindi tutti i regressori sono fortemente significativi nello spiegare .

L’ aumenta rispetto al caso precedente, anche se piccolo.

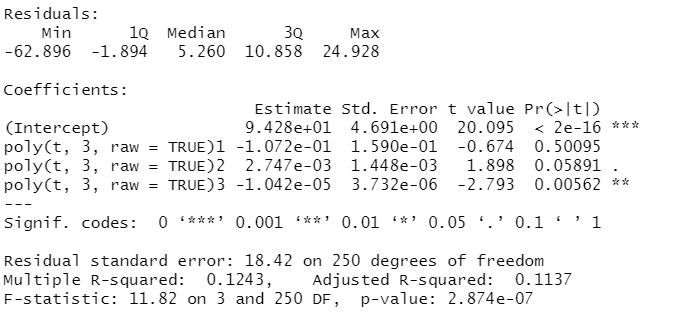
Immagine che contiene testo

Descrizione generata automaticamente



Confrontando i valori osservati e i valori stimati 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. Inoltre il valore β2<0 rende la curva rivolta verso il basso. Continuiamo nella stima del modello con ordine 3.



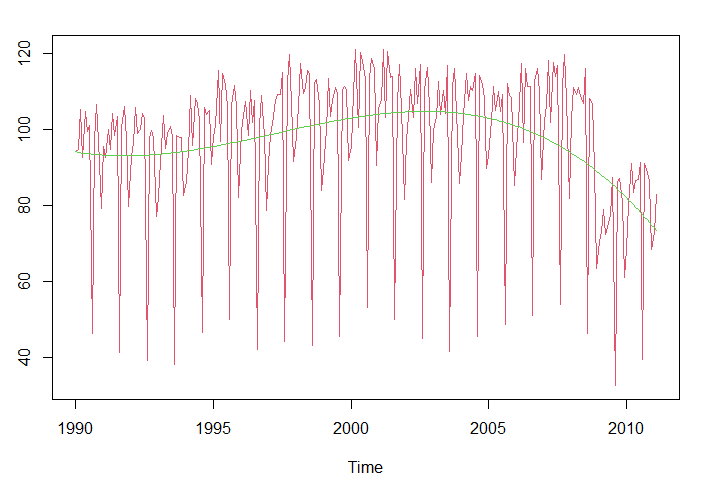


Il p-value del test di Fisher è minore di α, anche in questo caso rifiutiamo l’ipotesi nulla .

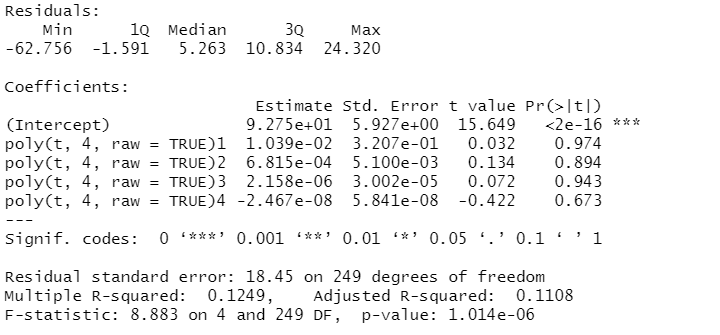
L’ aumenta lievemente rispetto al caso precedente, .

Immagine che contiene testo

Descrizione generata automaticamente



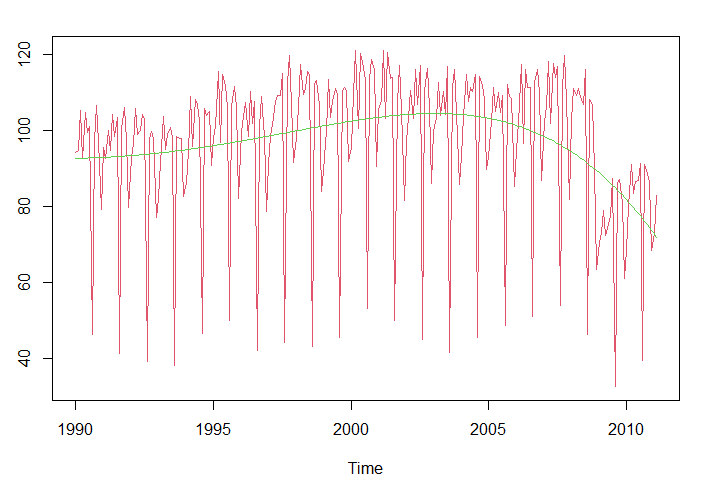
Il modello si adatta decisamente meglio, ma il criterio indica che si può proseguire stimando un modello di ordine 4.



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:

Per quanto riguarda il trend, concludiamo che il modello polinomiale di ordine 3 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.

Immagine che contiene tavolo

Descrizione generata automaticamenteProseguiamo stimando la stagionalità dalla serie storica utilizzando le variabili dummy. Nel nostro caso, essendo la serie mensile, abbiamo necessità di utilizzare 12 variabili dummy, ognuna delle quali indica il mese dell’anno in cui è stata osservata la serie.

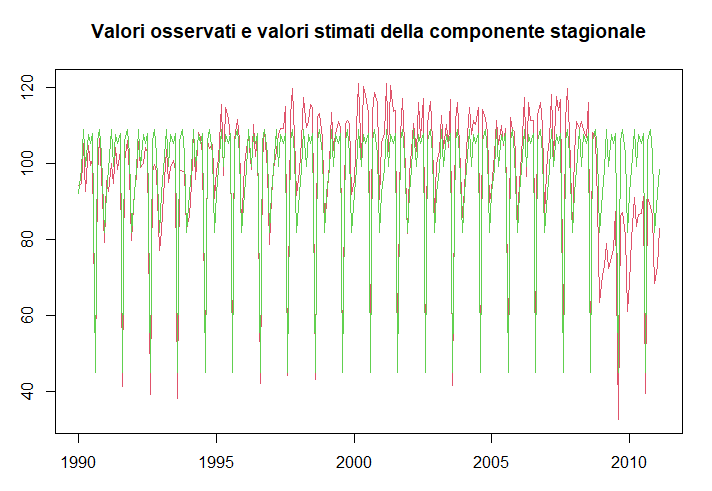
Costruiamo la stima ai minimi quadrati della componente stagionale:

Risulta opportuno inserire -1 nella specifica del modello per eliminare l’intercetta per evitare la trappola delle dummy e che essa comprometta l’adattabilità del modello

Immagine che contiene tavolo

Descrizione generata automaticamente

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

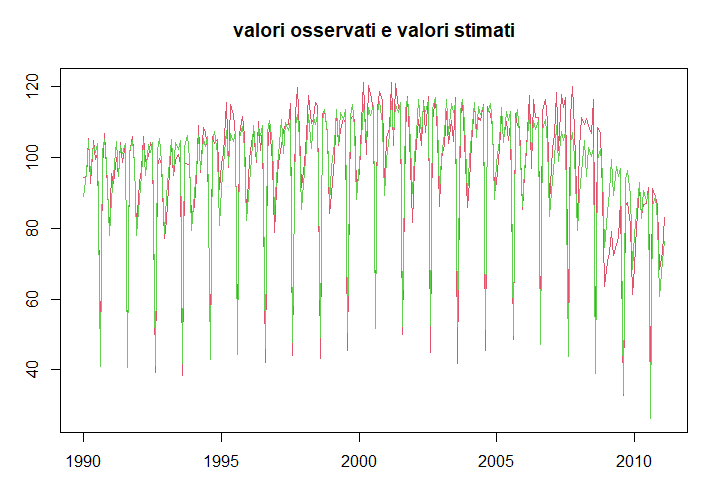


L'andamento della componente stagionale è più o meno conforme ai dati, studia bene la periodicità, ovviamente i nostri dati presentano anche la componente del trend.

Immagine che contiene tavolo

Descrizione generata automaticamenteStimiamo dunque il modello finale, che comprende sia la componente TREND che la componente stagionale:

Otteniamo un R2 molto alto pari a 0.9965 il che significa che il modello spiega quasi tutta la variabilità osservata. Confrontiamo a questo punto i dati con il modello finale:



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