

TOR VERGATA
UNIVERSITÀ DEGLI STUDI DI ROMA

RELAZIONE PROCESSI STOCASTICI ED ANALISI DI SERIE TEMPORALI

A.A. 2025/2026

ANDREA FERRETTI

0369448

Sommario

Simulazione processo white noise	3
Simulazione processo AR(1)	4
Processo MA(1).....	5
Trend e stagionalità	8
Dataset AirPassengers	14
Produzione industriale.....	17
Monthplot	18
Seasonal plot	19
Lag plot	20
Detrendizzazione	21
Detrendizzazione modello moltiplicativo	23
Vendita di alcol	25
Monthplot	27
Seasonal plot	27
Lag plot	29
Detrendizzazione e destagionalizzazione	29
Concentrazione CO2	32
Monthplot	34
Seasonal plot	34
Lag plot	35
Detrendizzazione e destagionalizzazione	37

Simulazione processo white noise

Un processo stocastico X_t si dice white noise se:

1. $E(X_t) = 0 \quad \forall t$
2. $Var(X_t) = E((X_t)^2) = \sigma^2 \quad \forall t, \sigma^2 \text{ costante}$
3. $R(t_1, t_2) = \sigma^2 \delta(t_2 - t_1) = \begin{cases} 0, & \text{se } t_1 \neq t_2 \\ \sigma^2, & \text{se } t_1 = t_2 \end{cases}$

Per simulare un processo di white noise in R si utilizza la funzione

```
> white_noise <- arima.sim(list(order=c(0,0,0)), n = 200)
```

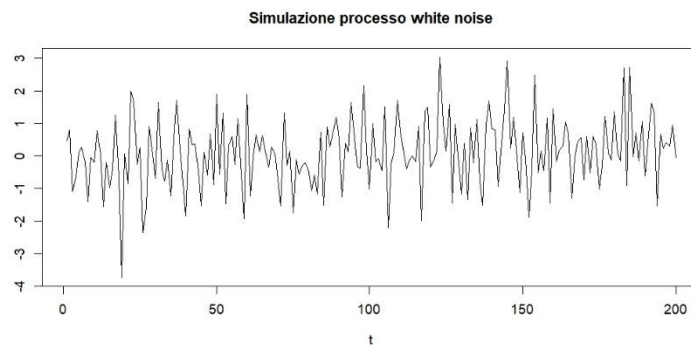
Gli argomenti passati in input sono:

- Order: indica l'ordine del processo ARIMA. c è il vettore che indica, rispettivamente, l'ordine p della componente autoregressiva, l'ordine d di differenziazione e l'ordine q della componente a media mobile.
- n : è il numero di realizzazioni del processo che si vogliono generare

È possibile visualizzare l'andamento del processo con l'istruzione

```
> plot(white_noise, ylab="", xlab="t", main="Simulazione processo white noise")
```

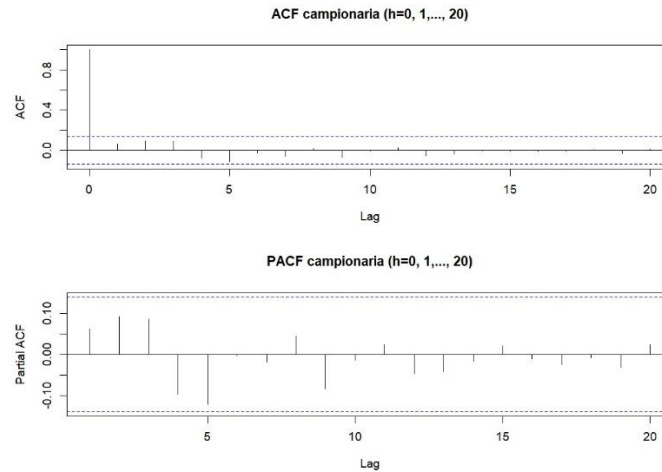
Il grafico ottenuto è il seguente:



Per calcolare e mostrare i grafici dell'autocorrelazione e dell'autocorrelazione parziale si utilizzano le seguenti funzioni:

```
> par(mfrow=c(2,1))
> acf(white_noise, lag.max=20, main="ACF campionaria (h=0, 1,..., 20)")
> pacf(white_noise, lag.max=20, main="PACF campionaria (h=0, 1,..., 20)")
```

La prima funzione serve per creare una tabella di grafici, composta da due righe e una colonna. *acf* calcola la funzione di autocorrelazione campionaria, mentre *pacf* calcola la funzione di autocorrelazione parziale campionaria. In entrambi i casi è stato scelto un ritardo massimo pari a venti. Di seguito sono riportati i grafici generati:



Nei due grafici è possibile notare la deviazione della acf e della pacf campionarie da quelle ideali: infatti, nel caso di un processo white noise la acf e la pacf teoriche dovrebbero essere

$$\rho(h) = \begin{cases} 1, & h = 0 \\ 0, & h \neq 0 \end{cases} \quad \varphi(k) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

uno per ritardo nullo, e zero per tutti gli altri; in questo caso, per ritardi non nulli entrambi i valori sono diversi da zero. Tuttavia, è possibile notare che solo una spike su venti esce dalle bande di confidenza.

Simulazione processo AR(1)

Un processo stocastico y_t è detto generato da un modello autoregressivo di ordine p se

$$y_t = \eta + \phi_1(y_{t-1} - \eta) + \dots + \phi_p(y_{t-p} - \eta) + u_t$$

Dove $(\eta, \phi_1, \dots, \phi_p)$ sono i parametri del modello, e u_t è un processo di white noise.

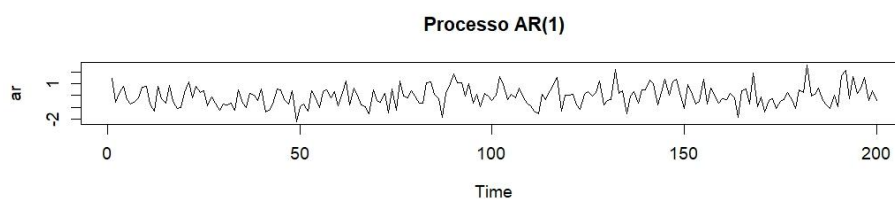
Si vuole simulare il seguente processo AR(1)

$$y_t = 0.2y_{t-1} + u_t$$

Per generare tale processo si utilizza il comando

```
> ar <- arima.sim(list(order=c(1, 0, 0), ar=0.2), n = 200)
```

Il parametro ar serve per passare alla funzione i parametri del modello.



Processo MA(1)

Un processo stocastico y_t è detto generato da un modello a media mobile di ordine q se

$$y_t = \eta + u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q}$$

Si vogliono simulare i seguenti processi MA(1):

$$Y_t = u_t + 0.5u_{t-1}$$

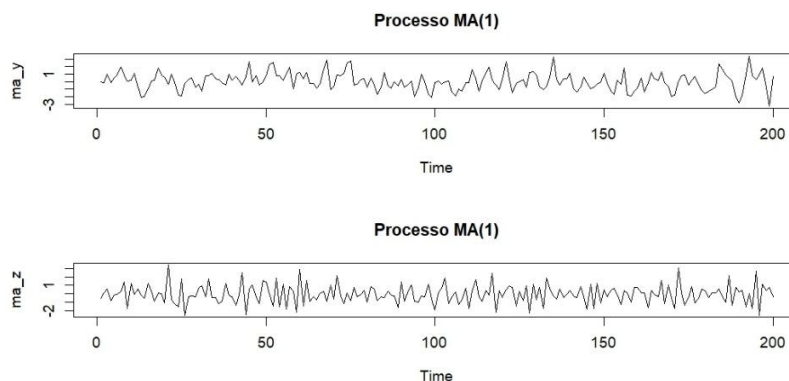
$$Z_t = u_t - 0.5u_{t-1}$$

A tale scopo, si eseguono le seguenti linee di codice:

```
> ma_y <- arima.sim(list(order=c(0,0,1), ma=0.5), n=200)
> ma_z <- arima.sim(list(order=c(0,0,1), ma=-0.5), n=200)
```

Successivamente, si mostrano i grafici eseguendo i comandi:

```
> par(mfrow=c(2,1))
> plot(ma_y, type="l", main="Processo MA(1)")
> plot(ma_z, type="l", main="Processo MA(1)")
```



Analizzando e confrontando i due grafici si evince che la serie $\{Z_t\}$ ha oscillazioni più frequenti della serie $\{Y_t\}$. Ciò può essere spiegato osservando le funzioni di autocorrelazione delle due serie. In particolare, si ha:

$$\rho_Y(h) = \begin{cases} 1, & h = 0 \\ \frac{0.5}{1 + 0.5^2} = 0.4, & h = 1 \\ 0, & h > 1 \end{cases} \quad \rho_Z(h) = \begin{cases} 1, & h = 0 \\ \frac{-0.5}{1 + (-0.5)^2} = -0.4, & h = 1 \\ 0, & h > 1 \end{cases}$$

Dalle due funzioni si nota che per ritardo $h = 1$ $\{Y_t\}$ ha autocorrelazione positiva, cioè a un aumento (o decremento) della serie tende ad esserci un aumento (o decremento) all'istante successivo. Per la serie $\{Z_t\}$ si ha autocorrelazione negativa per $h = 1$, cioè a un aumento (o decremento) della serie tende ad esserci un decremento (o aumento) della serie all'istante successivo.

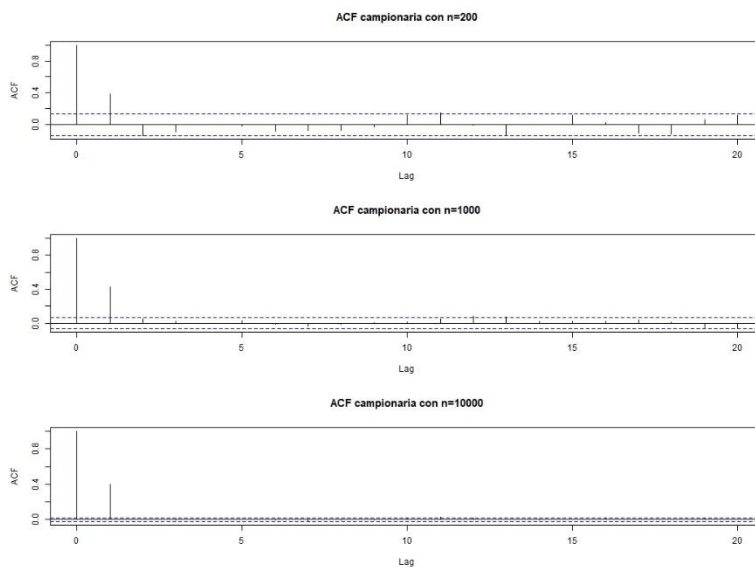
Stimiamo ora la funzione di autocorrelazione dei due processi al variare del numero dei campioni.

Per il processo $\{Y_t\}$ si generano le seguenti serie storiche:

```
> ma_y1 <- arima.sim(list(order=c(0,0,1), ma=0.5), n=200)
> ma_y2 <- arima.sim(list(order=c(0,0,1), ma=0.5), n=1000)
> ma_y3 <- arima.sim(list(order=c(0,0,1), ma=0.5), n=10000)
```

Si costruiscono, quindi, i relativi correlogrammi:

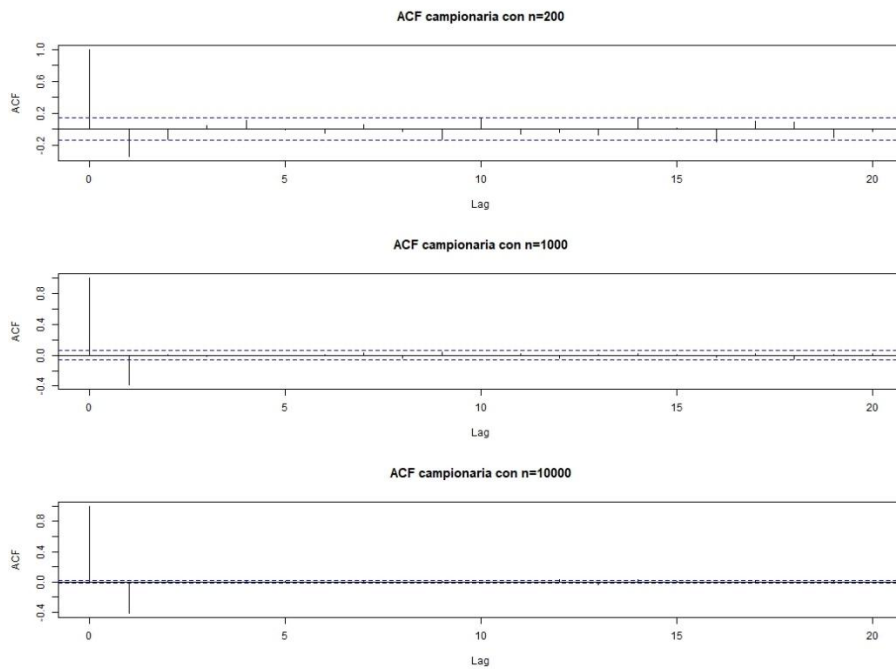
```
> par(mfrow=c(3,1))
> acf(ma_y1, lag.max=20, main="ACF campionaria con n=200")
> acf(ma_y2, lag.max=20, main="ACF campionaria con n=1000")
> acf(ma_y3, lag.max=20, main="ACF campionaria con n=10000")
```



Per il processo $\{Z_t\}$ i comandi da eseguire sono analoghi:

```
> ma_z1 <- arima.sim(list(order=c(0,0,1), ma=-0.5), n=200)
> ma_z2 <- arima.sim(list(order=c(0,0,1), ma=-0.5), n=1000)
> ma_z3 <- arima.sim(list(order=c(0,0,1), ma=-0.5), n=10000)
> par(mfrow=c(3,1))
> acf(ma_z1, lag.max=20, main="ACF campionaria con n=200")
> acf(ma_z2, lag.max=20, main="ACF campionaria con n=1000")
> acf(ma_z3, lag.max=20, main="ACF campionaria con n=10000")
```

Si ottengono i seguenti grafici:



È possibile notare che, nei grafici di entrambe le serie, la stima della funzione di autocorrelazione converge alla funzione di autocorrelazione teorica all'aumentare del numero delle realizzazioni delle serie temporali. Per verificare che i calcoli teorici siano corretti, è possibile eseguire le seguenti funzioni:

```
> exact_acf_y <- ARMAacf(ar=0, ma=0.5)
> exact_acf_z <- ARMAacf(ar=0, ma=-0.5)
```

I risultati ottenuti sono:

```
> exact_acf_y
```

0	1	2
1.0	0.4	0.0

```
> exact_acf_z
```

0	1	2
1.0	-0.4	0.0

I risultati confermano i calcoli teorici.

È possibile applicare la stessa funzione anche su altri processi MA, come ad esempio:

$$X_t = u_t + 0.2u_{t-1} - 0.15u_{t-2} + 0.2u_{t-3}$$

Il codice è analogo:

```
> exact_acf_x <- ARMAacf(ar=0, ma=c(0.2, -0.15, 0.2), lag.max=10)
> exact_acf_x
```

0	1	2	3	4	5	6	6	8	9	10
1.00000000	0.12698413	-0.09977324	0.18140590	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Trend e stagionalità

Se la serie temporale presenta un trend, è necessario detrendizzare la serie per renderla stazionaria e identificare il modello ARMA che la approssimi nel modo migliore.

Prima di eliminare il trend dalla serie temporale è necessario stimarlo: una tecnica che si può utilizzare è la stima attraverso la media mobile. In questo caso la stima del trend è:

$$\hat{T}_t = \frac{1}{2m+1} \sum_{k=-m}^m x_{t+k}$$

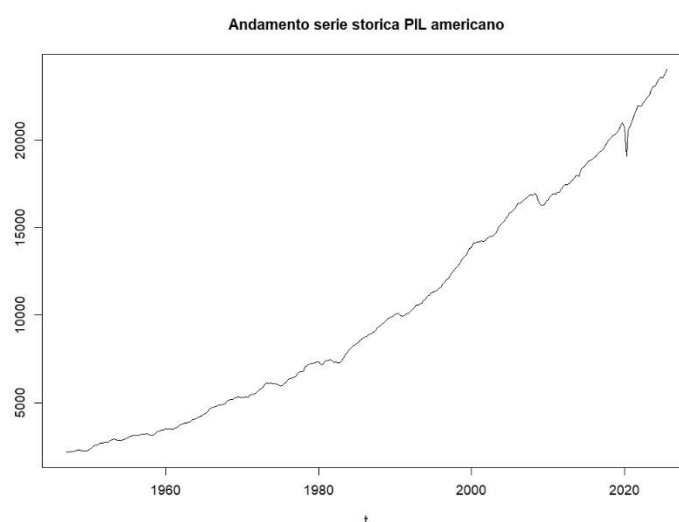
Se i dati hanno periodicità stagionale, la stima viene calcolata attraverso:

$$\hat{T}_t = \frac{1}{4m} (x_{t-m} + 2x_{t-m+1} + \dots + 2x_{t+m-1} + x_{t+m})$$

In particolare, per dati trimestrali si utilizza $m=2$, mentre per dati mensili si usa $m=6$.

Prendiamo la serie storica di dati trimestrali sul PIL americano dal 1947 fino ad oggi:

```
> pil.us <- read.csv("GDPC1.csv", sep=",")
> PIL <- pil.us$GDPC1
> xt <- ts(PIL, start=c(1947,1), freq=4)
> plot(xt, ylab="", xlab="t", main="Andamento serie storica PIL americano")
```



Dal grafico è possibile notare la presenza di un trend crescente. Per stimare il trend si utilizza il seguente vettore dei coefficienti:

$$\alpha = (\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$$

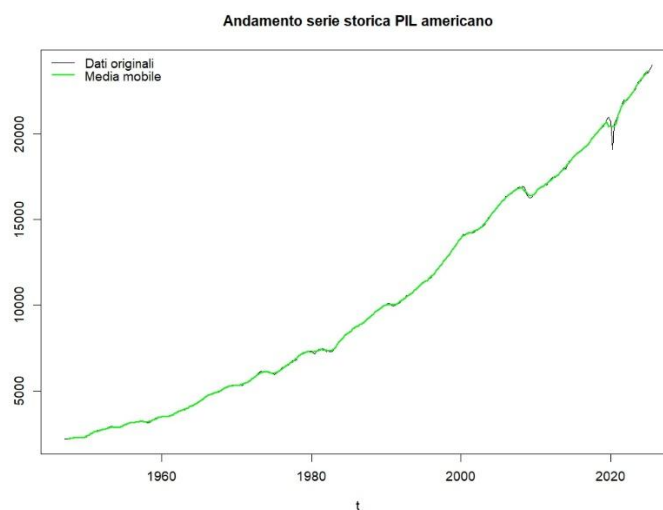
In tal caso si è posto $m=2$, essendo i dati trimestrali

Il vettore è inizializzato in R nel seguente modo:

```
> coefficients <- rep(1/5, 5)
```

Si calcola, quindi, la stima del trend:

```
> ma <- filter(xt, coefficients, method="convolution", sides=2, circular=FALSE)
> lines(ma, col="green", lwd=2)
> legend("topleft", legend=c("Dati originali", "Media mobile"), col=c("black",
"green"), lty=1, lwd=c(1,2), bty="n")
```



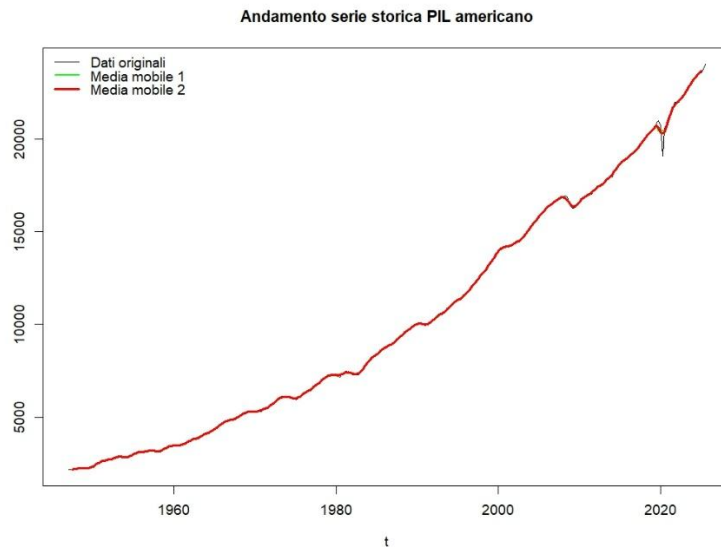
Gli argomenti della funzione `filter` sono: la serie di cui si vuole calcolare la media mobile (`xt`), il vettore dei coefficienti da utilizzare (`coefficients`), il metodo (“convolution”), le osservazioni utilizzate per calcolare il filtro (`sides`), formattazione dell’output (`circular`).

È possibile confrontare il trend stimato usando la formula generalizzata con la formula adatta per dati trimestrali. In tal caso la formula utilizzata è:

$$\hat{T}_t = \frac{1}{8}(x_{t-2} + 2x_{t-1} + 2x_t + 2x_{t+1} + x_{t+2})$$

Per calcolare questo filtro si eseguono i seguenti comandi:

```
> coefficients <- c(1/8, 1/4, 1/4, 1/4, 1/8)
> mac <- filter(xt, coefficients, method="convolution", sides=2, circular=FALSE)
> lines(mac, col="red", lwd=3)
> legend("topleft", legend=c("Dati originali", "Media mobile 1", "Media mobile 2"),
col=c("black", "green", "red"), lty=1, lwd=c(1,2,3), bty="n")
```

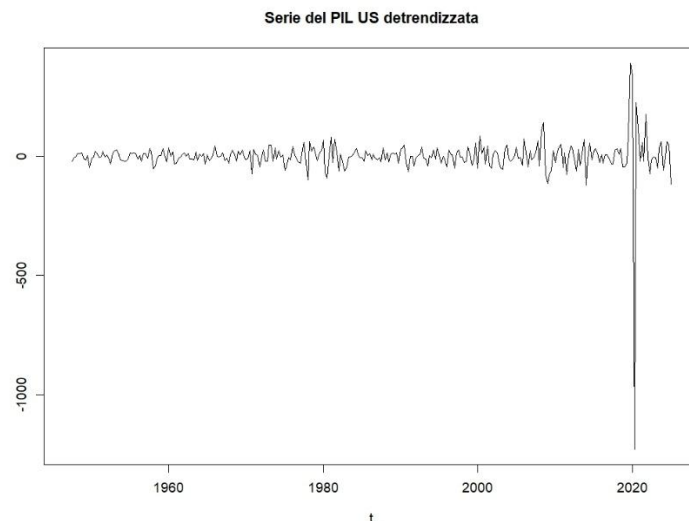


È possibile notare che non vi è molta differenza tra le due medie mobili. Se si ipotizza che la serie sia descritta mediante un modello additivo, è possibile detrendizzare la serie nel seguente modo:

$$\widehat{X}_t = X_t - \widehat{T}_t$$

Tale formula equivale in R a:

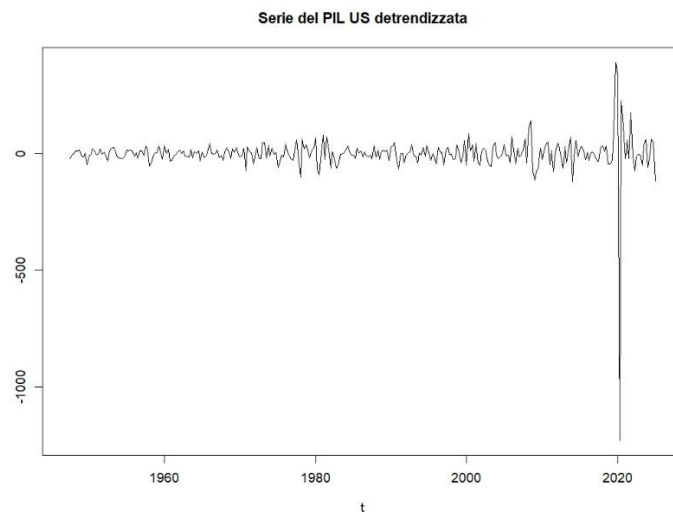
```
> detrendx <- xt - mac
> plot(detrendx, ylab="", xlab="t", main="Serie del PIL US detrendizzata")
```



Se, invece, si ipotizza che il modello è moltiplicativo, si detrendizza la serie calcolando:

$$\widehat{X}_t = \frac{X_t}{\widehat{T}_t}$$

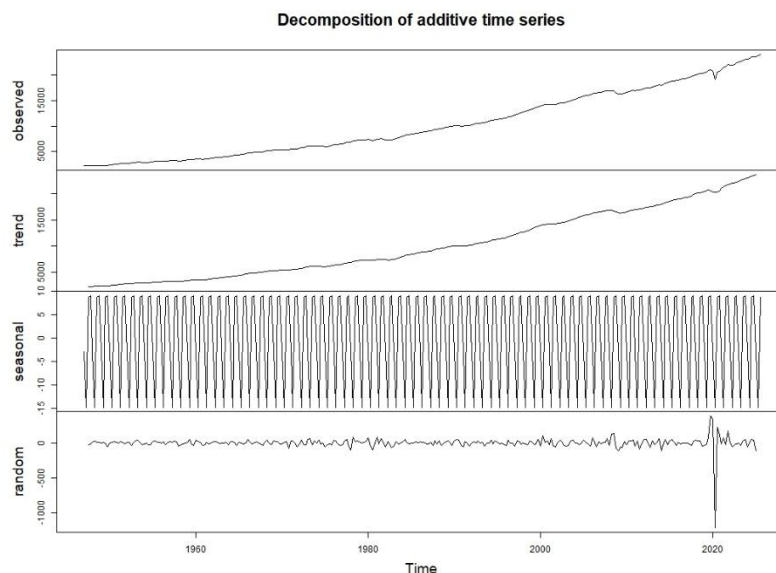
```
> detrendx <- xt / mac
> plot(detrendx, ylab="", xlab="t", main="Serie del PIL US detrendizzata")
```



Un altro modo per detrendizzare la serie temporale in R è attraverso la funzione **decompose()**. Questa funzione restituisce la stima del trend, salvata nel campo **trend**; la stima della stagionalità, salvata nel campo **seasonal**; l'errore residuo, salvato nel campo **random**. Nel caso di modello additivo, la componente di errore residua è calcolata come

$$\hat{E}_t = X_t - \hat{T}_t - \hat{S}_t$$

```
> d.pil <- decompose(xt, type="additive")
> plot(d.pil)
```

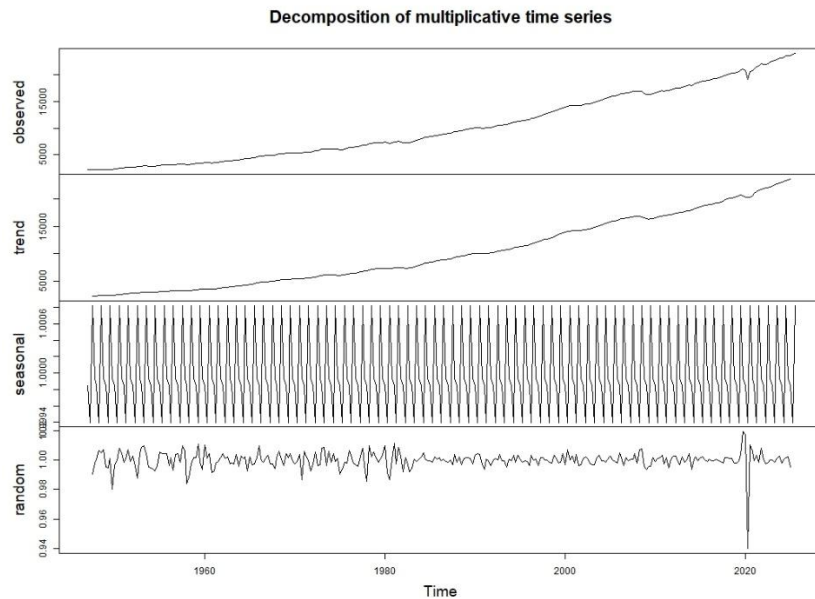


Nel caso di modello moltiplicativo si calcola

$$\hat{E}_t = \frac{X_t}{\hat{T}_t \hat{S}_t}$$

```
> d.pil <- decompose(xt, type="multiplicative")
```

```
> plot(d.pil)
```



Il terzo modo per detrendizzare la serie temporale consiste nell'applicare l'operatore differenza:

$$\nabla^d = (1 - B)^d$$

Per applicare l'operatore differenza, in R si usa:

```
> dxt <- diff(xt, lag=1, differences=1)
```

Dove x_t è la serie storica su cui si vuole applicare l'operatore differenza, lag è la potenza da applicare all'operatore ritardo, e differences è l'ordine d .

Nel caso della serie del PIL è stata calcolata la seguente serie:

$$dxt \equiv \nabla X_t = (1 - B)X_t = X_t - X_{t-1}$$

Se si vuole anche destagionalizzare la serie temporale, si può applicare il seguente operatore differenza:

$$\nabla_p^D = (1 - B^p)^D$$

Che in R corrisponde a usare la stessa funzione **diff()**, ma impostando opportunamente i parametri lag e differences . Essendo i dati trimestrali, si calcola:

```
> ddxt <- diff(dxt, lag=4, differences=1)
```

La serie detrendizzata e destagionalizzata è uguale a:

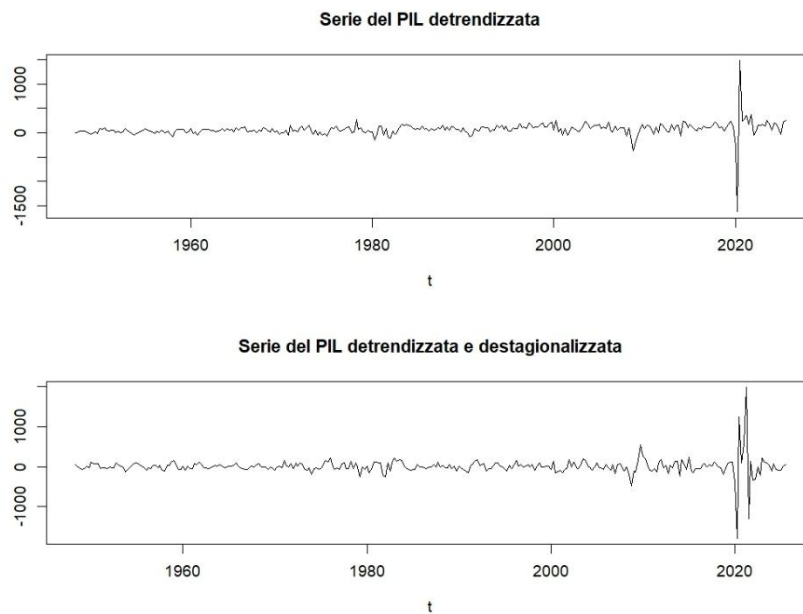
$$ddxt \equiv Y_t - Y_{t-4} = \nabla_4 \nabla X_t$$

Di seguito sono riportati i grafici delle serie dxt e $ddxt$:

```

> par(mfrow=c(2,1))
> plot(dxt, ylab="", xlab="t", main="Serie del PIL detrendizzata")
> plot(ddxt, ylab="", xlab="t", main="Serie del PIL detrendizzata e
destagionalizzata")

```

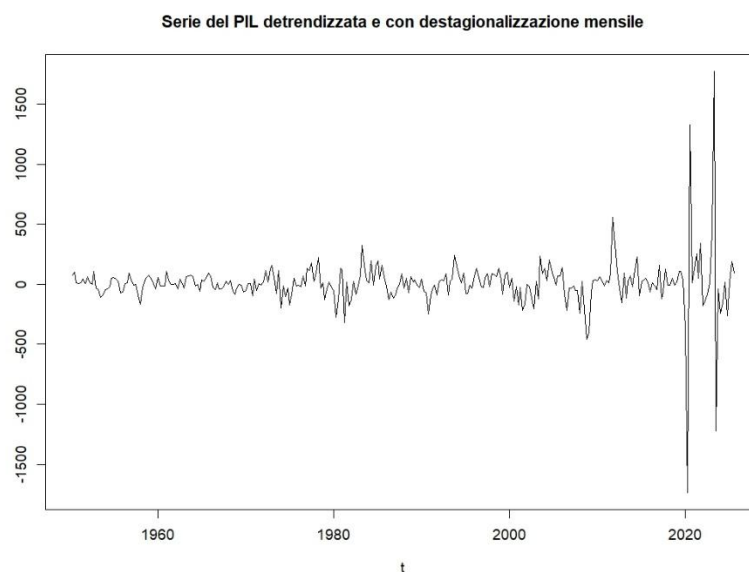


Se i dati fossero stati mensili, l'operatore da applicare sarebbe stato:

```

> ddxt.mensile <- diff(dxt, lag=12, differences=1)
> plot(ddxt.mensile, ylab="", xlab="t", main="Serie del PIL detrendizzata e con
destagionalizzazione mensile")

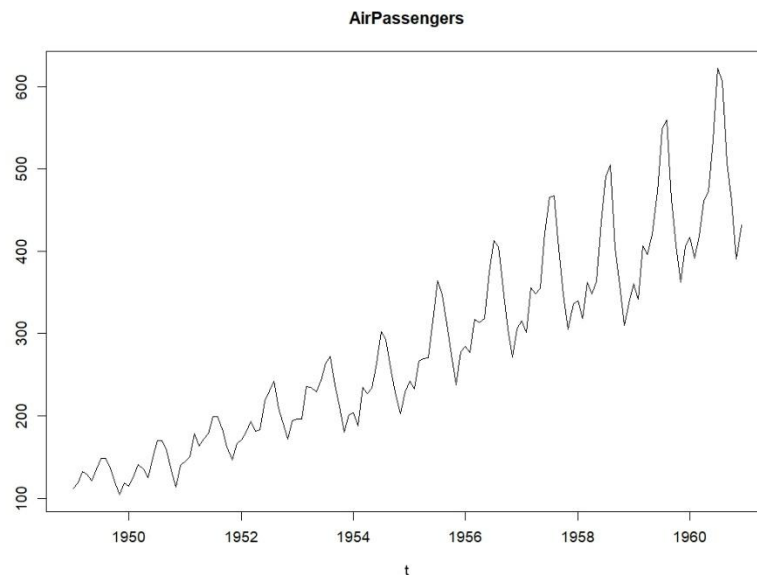
```



Dataset AirPassengers

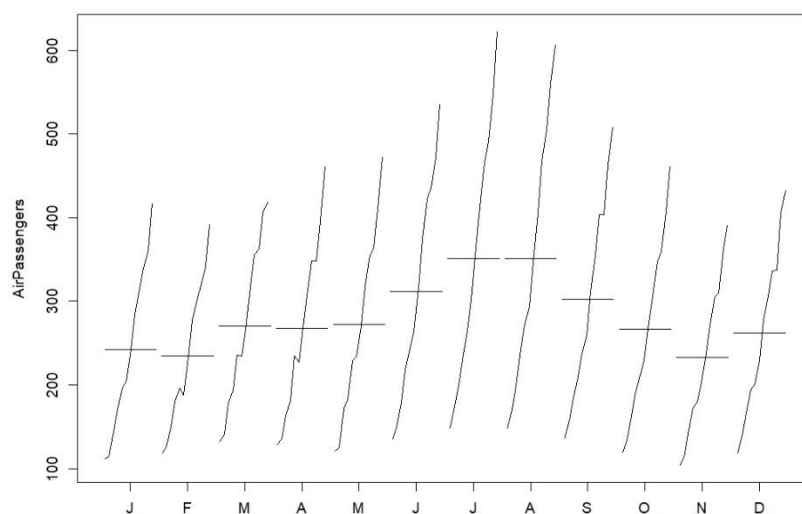
Consideriamo ora il dataset AirPassengers.

```
> data("AirPassengers")  
> plot(AirPassengers, ylab="", xlab="t", main="AirPassengers")
```



Dal grafico è possibile notare un trend crescente e una stagionalità. Per analizzare la stagionalità si esegue la funzione:

```
> monthplot(AirPassengers)
```

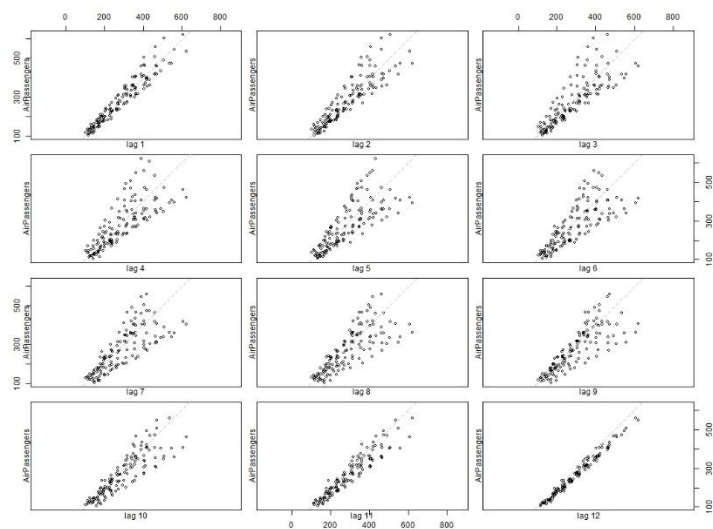


Le linee orizzontali rappresentano le medie delle sottoserie relative ai singoli mesi. Dal grafico è possibile notare un valore medio di passeggeri più alto nei mesi di Luglio e Agosto: questo ci

suggerisce di applicare l'operatore ∇_{12} . Le linee orizzontali indicano l'andamento delle sottoserie temporali: l'andamento crescente di tali linee ci suggerisce di utilizzare l'operatore ∇ .

Per individuare la presenza di trend e stagionalità si può utilizzare il diagramma di autodispersione, costituito dai punti $\{(X_t, X_{t+k}) : t = 1, 2, \dots, n - k\}$, con k che varia.

```
> lag.plot(AirPassengers, set.lags=1:12, type="p", do.lines=FALSE)
```



Set.lags indica i valori di ritardo k per cui si vogliono calcolare diagrammi di autodispersione. Dai diagrammi si nota la presenza di un trend lineare crescente, poiché i punti si disperdono lungo la retta tratteggiata. Inoltre, per $k=12$, cioè stesso mese, ma anno precedente, i valori sono molto simili tra di loro: ciò conferma la presenza di stagionalità.

Per analizzare meglio la stagionalità della serie storica si ricorre al seasonal plot, che mostra l'andamento della serie storica, limitata ad un singolo anno. Ad esempio, è possibile visualizzare l'andamento della serie storica negli anni 1949, 1951, 1954, 1956, 1958 e 1960 nel seguente modo:

```
> y1 <- window(AirPassengers, start=c(1949,1), end=c(1949,12))
> y2 <- window(AirPassengers, start=c(1951,1), end=c(1951,12))
> y3 <- window(AirPassengers, start=c(1954,1), end=c(1954,12))
> y4 <- window(AirPassengers, start=c(1956,1), end=c(1956,12))
> y5 <- window(AirPassengers, start=c(1958,1), end=c(1958,12))
> y6 <- window(AirPassengers, start=c(1960,1), end=c(1960,12))
```

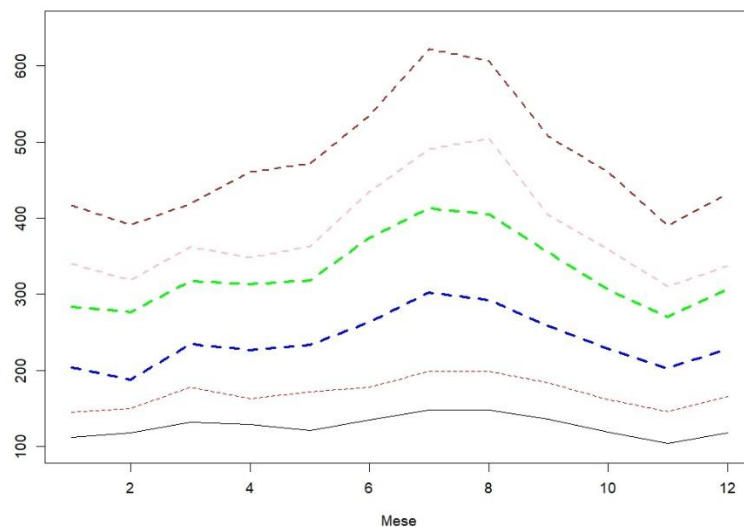
Queste serie contengono 12 osservazioni ciascuna, una per ogni mese.

```
> range(AirPassengers)
```

Questa istruzione serve per ottenere il valore massimo e minimo della serie storica originale. Il risultato restituito è [1] 104 622

```
> x <- seq(1, 12, 1)
> matplot(x, y1, type="l", ylim=c(100,650), ylab="", xlab="Mese")
> lines(x, y2, type="l", lty=2, col="red")
> lines(x, y3, type="l", lty=2, col="blue", lwd=3)
> lines(x, y4, type="l", lty=2, col="green", lwd=3)
> lines(x, y5, type="l", lty=2, col="pink", lwd=2)
> lines(x, y6, type="l", lty=2, col="brown", lwd=2)
```

La prima istruzione genera i valori lungo l'asse delle ascisse, mentre le linee rimanenti generano il grafico, che viene riportato di seguito:



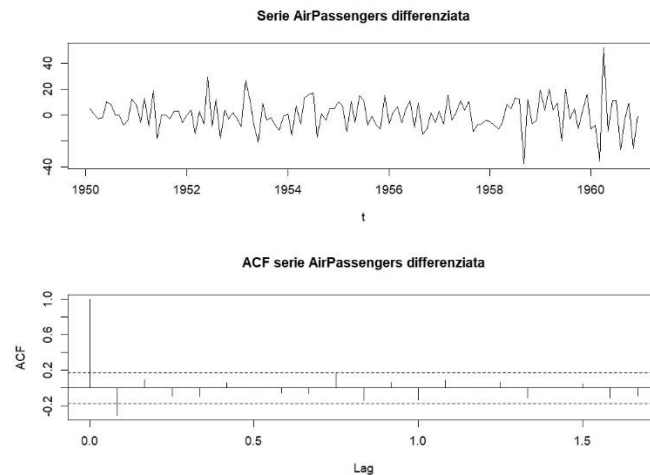
Dal grafico si evince che i valori della serie nel corso degli anni aumentano in corrispondenza di tutti i mesi. Inoltre, le linee delle sottoserie storiche hanno forme molto simili tra di loro, con un picco nei mesi di Luglio e Agosto, confermando così la stagionalità della serie storica, e la presenza di un trend crescente.

Dunque, si applicano gli operatori ∇ e ∇_{12} , utilizzati rispettivamente per detrendizzare e destagionalizzare la serie storica.

```
> yt <- diff(AirPassengers, lag=1, differences=1)
> zt <- diff(yt, lag=12, differences=1)
```



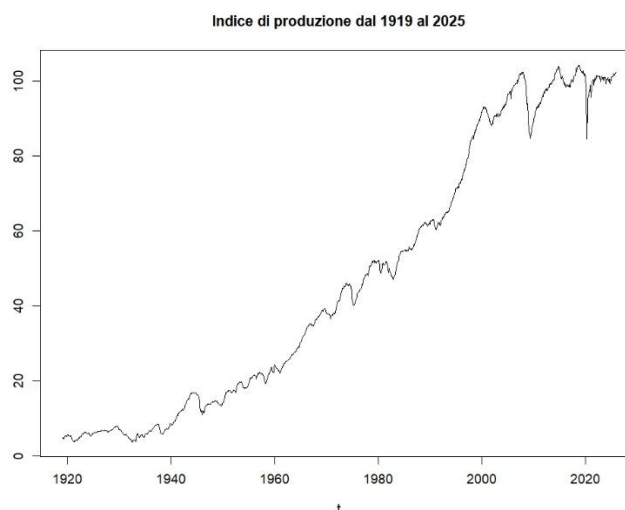
```
> par(mfrow=c(2,1))
> plot(zt, ylab="", xlab="t", main="Serie AirPassengers differenziata")
> acf(zt, lag.max=20, main="ACF serie AirPassengers differenziata")
```



Produzione industriale

Il dataset è stato scaricato dal [sito](#) della Federal Reserve, e contiene l'andamento mensile dell'indice di produzione dei settori manifatturiero, minerario ed energetico degli Stati Uniti. Per calcolare l'indice è stato utilizzato come riferimento il 2017.

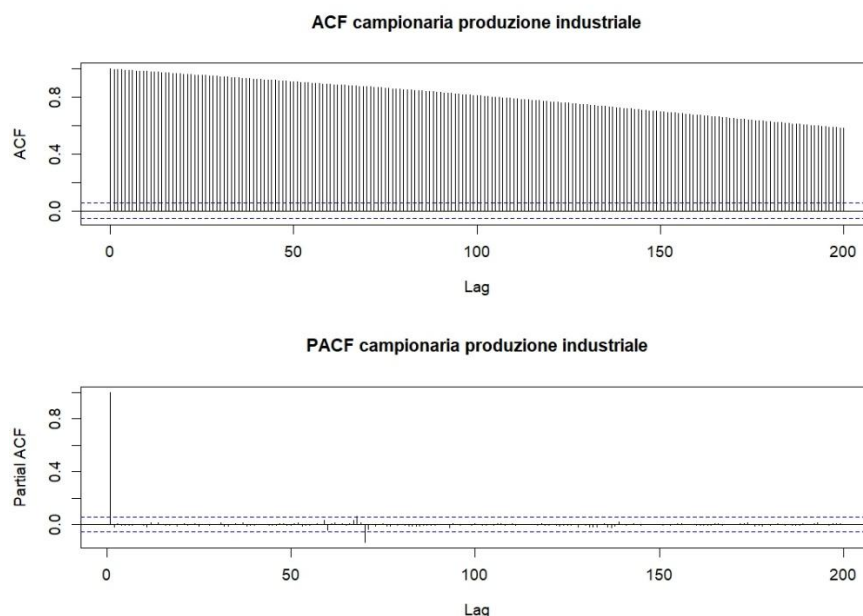
```
> ind_prod <- read.csv("./INDPRO.csv", sep=",")
> ind_prod_values <- ind_prod$INDPRO
> dates <- as.Date(ind_prod$observation_date, format="%Y-%m-%d")
> plot(dates, ind_prod_values, type="l", xlab="t", ylab="", main="Indice di
produzione dal 1919 al 2025")
```



Dal grafico è possibile notare che vi è un trend crescente, con una stabilizzazione a partire dal primo decennio degli anni 2000: questo rallentamento della produzione industriale potrebbe essere dovuto alla crisi del 2008 e al covid nel 2020, corrispondenti ai picchi negativi; ciò ha fatto sì che la produzione industriale dovesse ritornare al regime degli anni precedenti prima di poter iniziare a crescere di nuovo.

Per prima cosa, si analizza l'autocorrelazione della serie temporale:

```
> par(mfrow=c(2,1))  
> acf(ind_prod_values, lag.max=200, main="ACF campionaria produzione industriale")  
> pacf(ind_prod_values, lag.max=200, main="PACF campionaria produzione industriale")
```



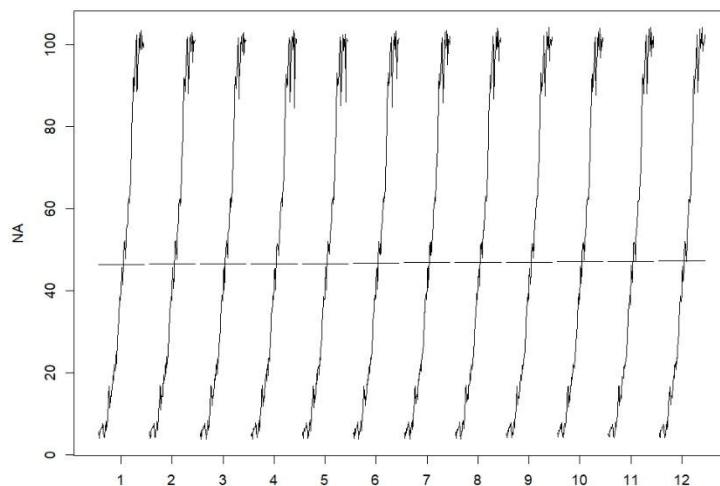
Il grafico della acf campionaria mostra un decadimento lineare della funzione, confermando la non stazionarietà della serie temporale, e suggerendo la presenza di un trend.

Dal grafico della pacf campionaria si evince che vi è una autocorrelazione parziale significativa solo per lag uguale a 1: ciò suggerisce, da un lato, che la serie temporale non abbia alcuna componente stagionale; dall'altro, avvalora l'ipotesi della presenza di un trend.

Monthplot

Per analizzare meglio il trend della serie temporale, si utilizza il monthplot:

```
> monthplot(ind_prod_values)
```



Anche questo grafico conferma una crescita lineare per ciascun mese dell'anno. Inoltre, i valori medi sui singoli mesi sono molto simili tra di loro, per cui non sembra esserci alcun tipo di stagionalità.

Seasonal plot

Per verificare ulteriormente l'assenza di stagionalità, si ricorre al seasonal plot:

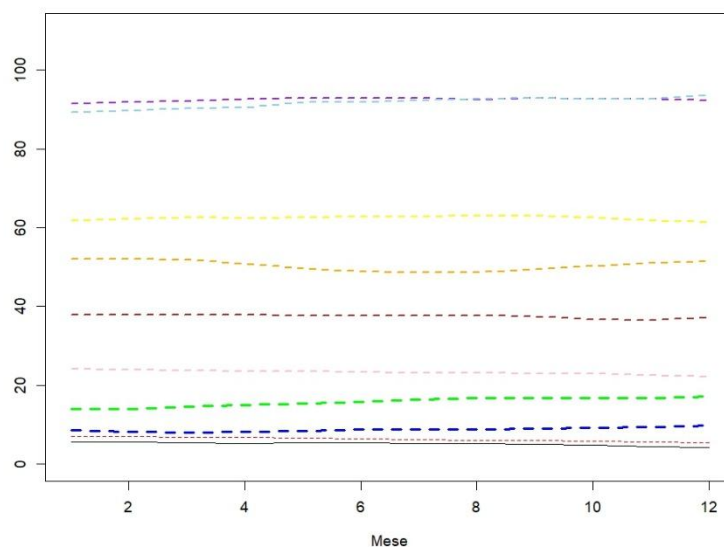
```
> ind_prod_ts <- ts(ind_prod_values, start=c(1919,1), frequency=12)
> y1 <- window(ind_prod_ts, start=c(1920,1), end=c(1920,12))
> y2 <- window(ind_prod_ts, start=c(1930,1), end=c(1930,12))
> y3 <- window(ind_prod_ts, start=c(1940,1), end=c(1940,12))
> y4 <- window(ind_prod_ts, start=c(1950,1), end=c(1950,12))
> y5 <- window(ind_prod_ts, start=c(1960,1), end=c(1960,12))
> y6 <- window(ind_prod_ts, start=c(1970,1), end=c(1970,12))
> y7 <- window(ind_prod_ts, start=c(1980,1), end=c(1980,12))
> y8 <- window(ind_prod_ts, start=c(1990,1), end=c(1990,12))
> y9 <- window(ind_prod_ts, start=c(2000,1), end=c(2000,12))
> y10 <- window(ind_prod_ts, start=c(2010,1), end=c(2010,12))
```

La prima riga di codice serve per poter trasformare il vettore `ind_prod_values` in una serie temporale, in modo tale da poter creare delle finestre temporali.

```
> range(ind_prod_ts)
```

[1] 3.6891 104.1004

```
> x <- seq(1, 12, 1)
> matplot(x, y1, type="l", ylim=c(0,110), ylab="", xlab="Mese")
> lines(x, y2, type="l", lty=2, col="red")
> lines(x, y3, type="l", lty=2, col="blue", lwd=3)
> lines(x, y4, type="l", lty=2, col="green", lwd=3)
> lines(x, y5, type="l", lty=2, col="pink", lwd=2)
> lines(x, y6, type="l", lty=2, col="brown", lwd=2)
> lines(x, y7, type="l", lty=2, col="orange", lwd=2)
> lines(x, y8, type="l", lty=2, col="yellow", lwd=2)
> lines(x, y9, type="l", lty=2, col="purple", lwd=2)
> lines(x, y10, type="l", lty=2, col="skyblue", lwd=2)
```

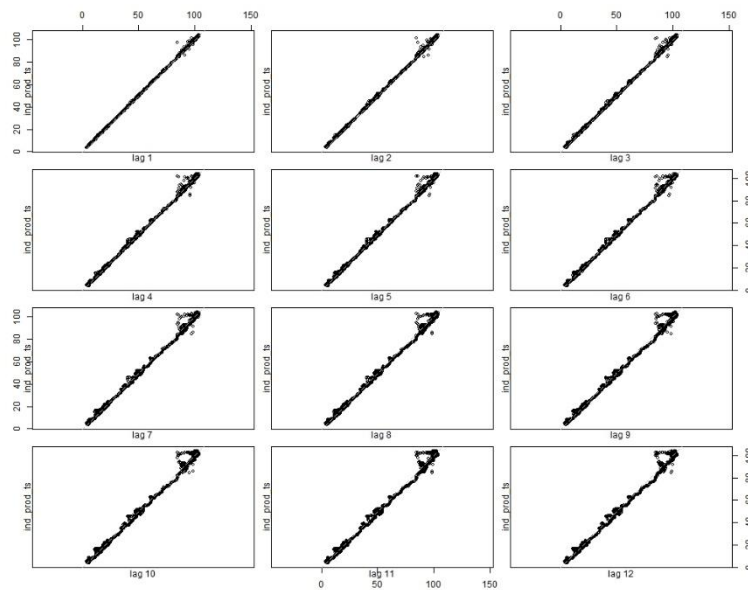


Dal grafico si nota che non vi è alcuna presenza di stagionalità. Oltre a ciò, emerge che nei primi decenni l'aumento di produzione non è stato molto alto, probabilmente per via della crisi del 1929 e della seconda guerra mondiale; successivamente, si ha un aumento costante della produzione, che culmina con un picco tra il 1990 e il 2000, per poi arrestarsi tra il 2000 e il 2010 a causa della crisi del 2008.

Lag plot

Per avere un quadro completo della serie temporale, si studia il lag plot:

```
> lag.plot(ind_prod_ts, set.lags=1:12, type="p", do.lines=FALSE)
```



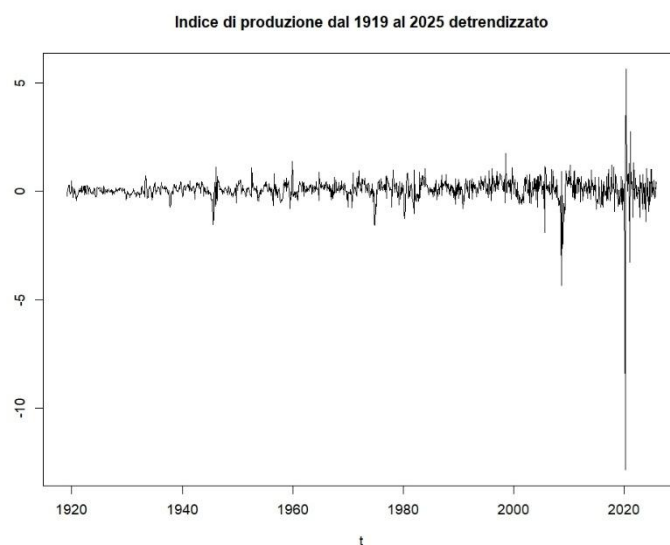
Dal lag plot non è possibile concludere molto sulla stagionalità della serie temporale, poiché il trend lineare nasconde eventuali pattern che si ripetono a intervalli temporali costanti. Si decide, quindi, di detrendizzare la serie temporale, e di analizzare il lag plot della serie temporale ottenuta.

Detrendizzazione

Ipotizzando che il trend sia lineare, si calcola:

$$Z_t = \nabla X_t = (1 - B)X_t$$

```
> detr_ind_prod_values <- diff(ind_prod_values, lag=1, diff=1)
> plot(dates[-1], detr_ind_prod_values, type="l", xlab="t", ylab="", main="Indice di
produzione dal 1919 al 2025 detrendizzato")
```

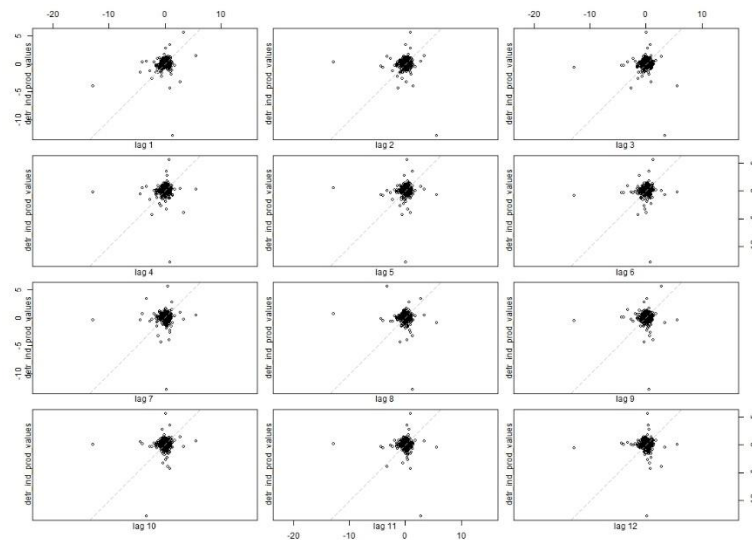


Il -1 usato in dates serve per eliminare il primo elemento della serie, il quale non ha una realizzazione che lo precede.

Dal grafico si può notare che la serie temporale detrendizzata sembra avere media nulla.

Si analizza ora il nuovo lag plot ottenuto:

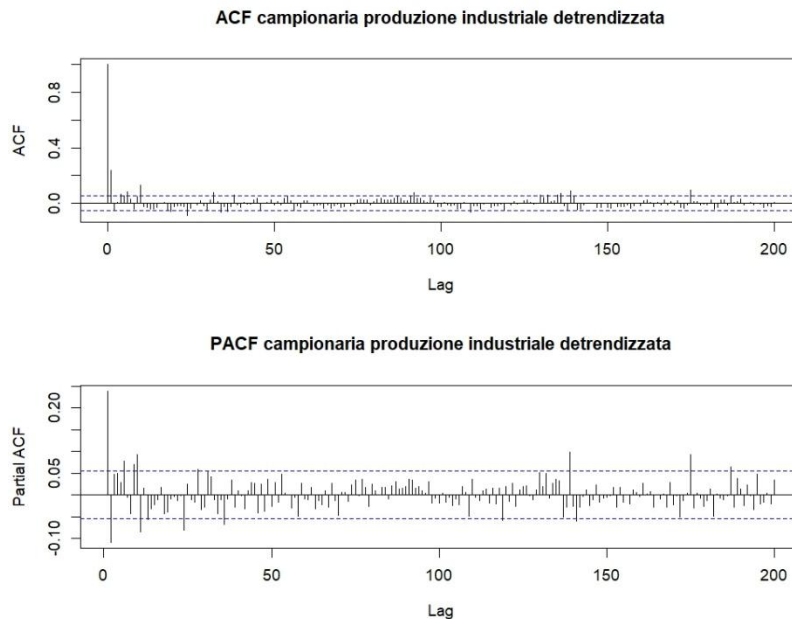
```
> par(mfrow=c(2,1))  
  
> acf(detr_ind_prod_values, lag.max=200, main="ACF campionaria produzione industriale  
detrendizzata")  
  
> pacf(detr_ind_prod_values, lag.max=200, main="PACF campionaria produzione industriale  
detrendizzata")
```



Il lag plot conferma che la serie temporale non presenta componenti stagionali, poichè la distribuzione dei punti è simile per ogni lag. Inoltre, si nota che i punti si distribuiscono intorno allo 0, in accordo con la media nulla della serie temporale detrendizzata, con degli outlier che si disperdono molto, probabilmente dovuti alle crisi economiche dell'ultimo secolo.

Di seguito è riportato il correlogramma:

```
> par(mfrow=c(2,1))  
  
> acf(detr_ind_prod_values, lag.max=200, main="ACF campionaria produzione industriale  
detrendizzata")  
  
> pacf(detr_ind_prod_values, lag.max=200, main="PACF campionaria produzione  
industriale detrendizzata")
```



La acf campionaria riporta correlazioni significative per diversi lag, con un numero non trascurabile di valori al di fuori delle bande di confidenza: ciò suggerisce che la serie detrendizzata non sia stazionaria.

Per verificare se la serie temporale detrendizzata è stazionaria, si esegue il test di Ljung-Box:

```
> Box.test(detr_ind_prod_values, lag=20, type="Ljung-Box")
```

Box-Ljung test

data: detr_ind_prod_values

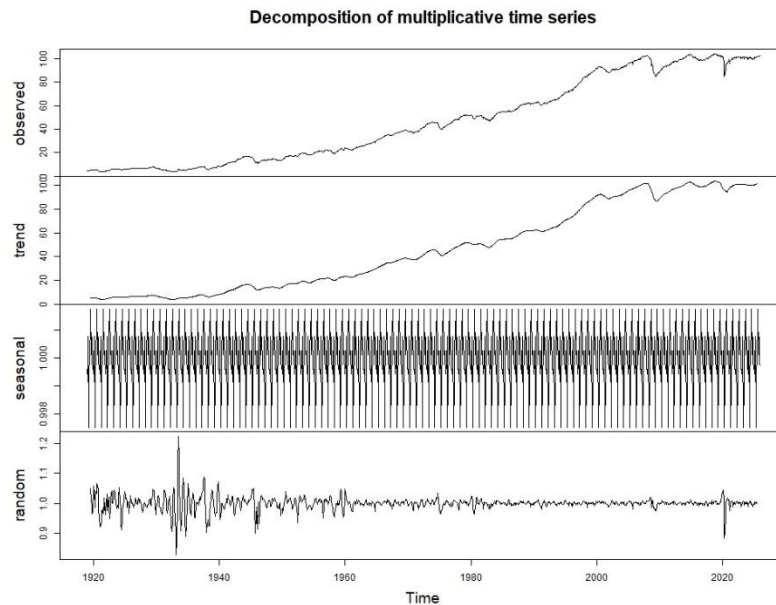
X-squared = 138.3, df = 20, p-value = 1.203e-11

Il test restituisce un p-value molto minore di 0.05, pertanto si decide di rifiutare l'ipotesi nulla che la serie temporale detrendizzata sia un white noise.

Detrendizzazione modello moltiplicativo

Per studiare meglio la serie temporale, si procede con il trattare la serie temporale utilizzando un modello moltiplicativo:

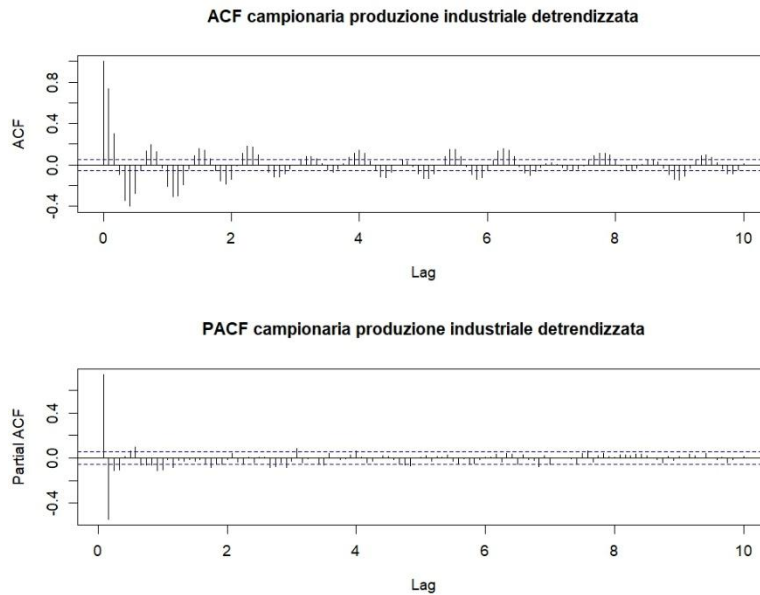
```
> detr_ind_prod_values <- decompose(ind_prod_ts, type="multiplicative")
> plot(detr_ind_prod_values)
```



Il grafico dei residui ha un andamento speculare a quello della serie temporale detrendizzata con il modello additivo. È possibile notare una varianza molto alta per i primi decenni del Novecento, seguita da una diminuzione e una stabilizzazione per gli anni che vanno fino al 2000, e da picchi negativi in corrispondenza della crisi economica del 2008 e del COVID nel 2020.

Si riportano di seguito i correlogrammi:

```
> serie_pulita <- na.omit(detr_ind_prod_values$random)
> par(mfrow=c(2,1))
> acf(serie_pulita, lag.max=120, main="ACF campionaria produzione industriale
detrendizzata")
> pacf(serie_pulita, lag.max=120, main="PACF campionaria produzione industriale
detrendizzata")
```

Dai grafici notiamo che neanche una decomposizione effettuata utilizzando un modello moltiplicativo sia in grado di rendere la serie stazionaria. Per confermare l'analisi dei grafici, si riesegue il test di Ljung-Box:

```
> Box.test(serie_pulita, lag=20, type="Ljung-Box")
```

Box-Ljung test

data: serie_pulita

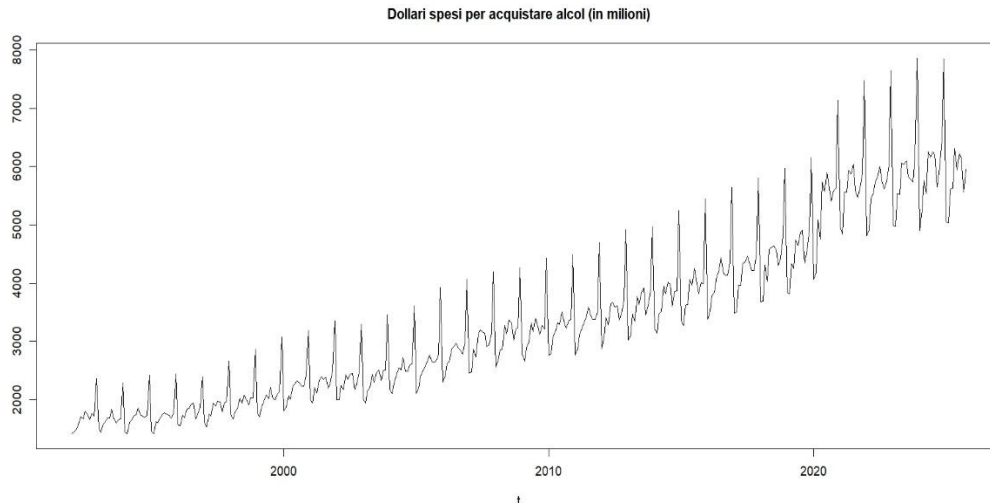
X-squared = 1802.6, df = 20, p-value < 2.2e-16

Il test rifiuta l'ipotesi nulla: il motivo è che ci sono componenti autoregressive o a media mobile che non sono state eliminate. Ciò è dovuto principalmente alle crisi del 1929, del 2008 e del 2020, anomalie che aumentano la varianza all'interno della serie temporale, e che non possono essere catturate dal trend.

Vendita di alcol

Il dataset, scaricato dal [sito FRED](#), indica la quantità di dollari (espressa in milioni) spesi ogni mese per acquistare birra, vino e liquori dal 1/1/1992 al 1/10/2025.

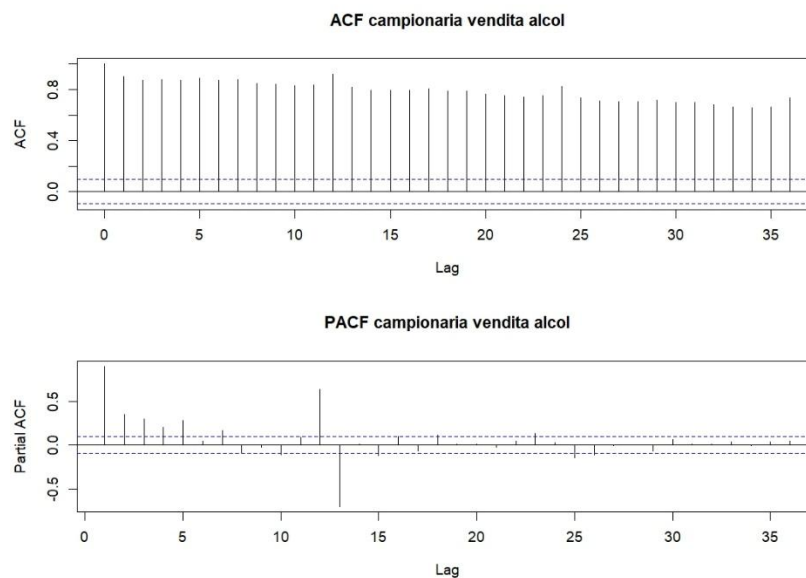
```
> alcol <- read.csv("./dataset.csv", sep=",")
> dates <- as.Date(alcol$observation_date, format="%Y-%m-%d")
> dollars <- alcol$Spesa
> plot(dates, dollars, type="l", ylab="", xlab="t", main="Dollari spesi per
acquistare alcol (in milioni)")
```



Dal grafico è possibile notare un trend crescente della serie temporale. Inoltre, si nota come la serie temporale abbia andamenti che si ripetono a intervalli regolari, per cui si ipotizza la presenza di una componente stagionale.

Si analizzano ora i correlogrammi della serie temporale:

```
> par(mfrow=c(2,1))
> acf(dollars, lag.max=36, main="ACF campionaria vendita alcol")
> pacf(dollars, lag.max=36, main="PACF campionaria vendita alcol")
```



Dal grafico della acf campionaria si nota un decremento lineare dei valori di autocorrelazione, i quali sono un chiaro segno di non stazionarietà.

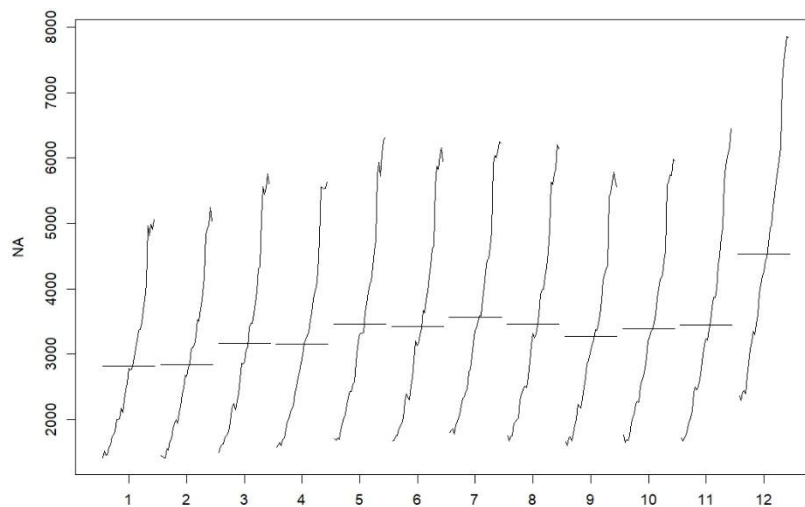
Il grafico della pacf campionaria mostra un picco al lag 1, dovuto al trend, oltre a dei valori significativi fino dal lag 2 al lag 5, e per i lag 12 e 13. Il picco in corrispondenza del lag 12 è un

ulteriore indizio della stagionalità, che si suppone essere annuale. Il picco in corrispondenza del lag 13 indica che la serie temporale tende ad avere un andamento opposto rispetto a quello avuto il mese precedente dell'anno precedente. I picchi tra il lag 2 e il lag 5 potrebbero indicare una certa inerzia della serie temporale.

Monthplot

Per analizzare il trend e la stagionalità della serie temporale, si utilizza il monthplot:

```
> monthplot(dollars)
```



Dal monthplot si nota la crescita lineare della quantità di dollari spesa in alcol. Inoltre, i valori mensili medi sono diversi tra di loro. In particolare, vi è una media più alta in corrispondenza di dicembre, probabilmente dovuta alle feste di fine anno, mentre le medie più basse si hanno a gennaio e febbraio. Negli altri mesi la medie assumono valori in un intervallo ristretto.

Seasonal plot

Per disegnare il seasonal plot si converte il vettore di elementi in una serie temporale:

```
> dollars_ts <- ts(dollars, start=c(1992,1), frequency=12)
```

Successivamente si creano le finestre temporali: la prima corrisponde all'anno 1992, mentre le rimanenti sono scelte con cadenza pari a cinque anni.

```
> y1 <- window(dollars_ts, start=c(1992,1), end=c(1992,12))  
> y2 <- window(dollars_ts, start=c(1997,1), end=c(1997,12))  
> y3 <- window(dollars_ts, start=c(2002,1), end=c(2002,12))
```

```

> y4 <- window(dollars_ts, start=c(2007,1), end=c(2007,12))
> y5 <- window(dollars_ts, start=c(2012,1), end=c(2012,12))
> y6 <- window(dollars_ts, start=c(2017,1), end=c(2017,12))
> y7 <- window(dollars_ts, start=c(2022,1), end=c(2022,12))

> range(dollars_ts)

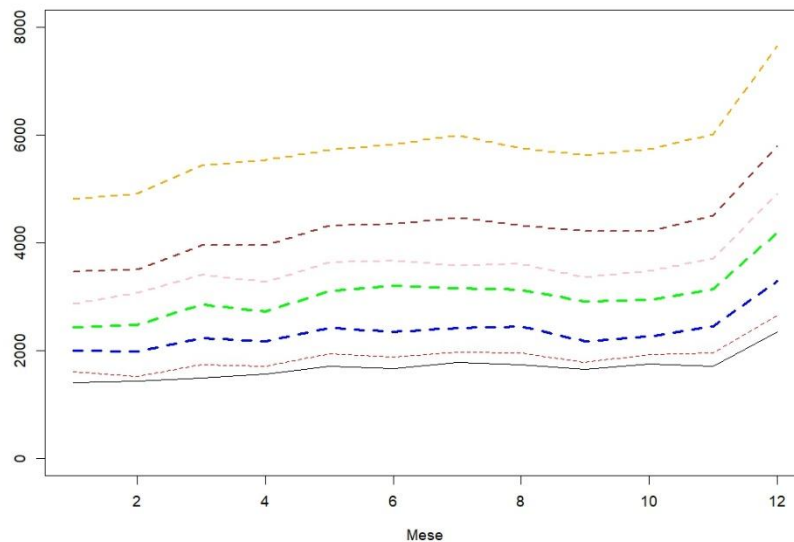
```

[1] 1407 7863

```

> x <- seq(1, 12, 1)
> matplot(x, y1, type="l", ylim=c(0,8000), ylab="", xlab="Mese")
> lines(x, y2, type="l", lty=2, col="red")
> lines(x, y3, type="l", lty=2, col="blue", lwd=3)
> lines(x, y4, type="l", lty=2, col="green", lwd=3)
> lines(x, y5, type="l", lty=2, col="pink", lwd=2)
> lines(x, y6, type="l", lty=2, col="brown", lwd=2)
> lines(x, y7, type="l", lty=2, col="orange", lwd=2)

```

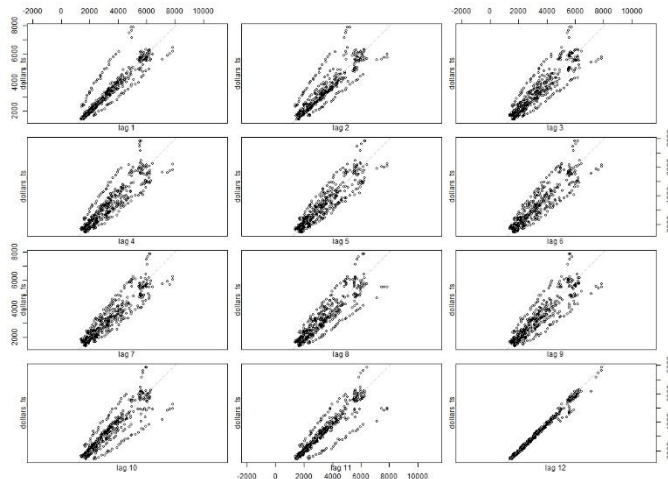


Dal grafico è possibile notare che le serie storiche annuali hanno andamento molto simile. Inoltre, la quantità di dollari spesi aumenta in modo costante nelle prime sei serie storiche, mentre nell'ultima vi è un aumento più marcato.

Lag plot

Per completezza, prima di detrendizzare e destagionalizzare la serie temporale, si mostra anche il lag plot.

```
> lag.plot(dollars_ts, set.lags=1:12, type="p", do.lines=FALSE)
```



Il grafico relativo al lag pari a 12 conferma la stagionalità annuale: i punti, infatti, sono disposti lungo una linea retta, con qualche punto che si allontana da essa, anche se non di molto.

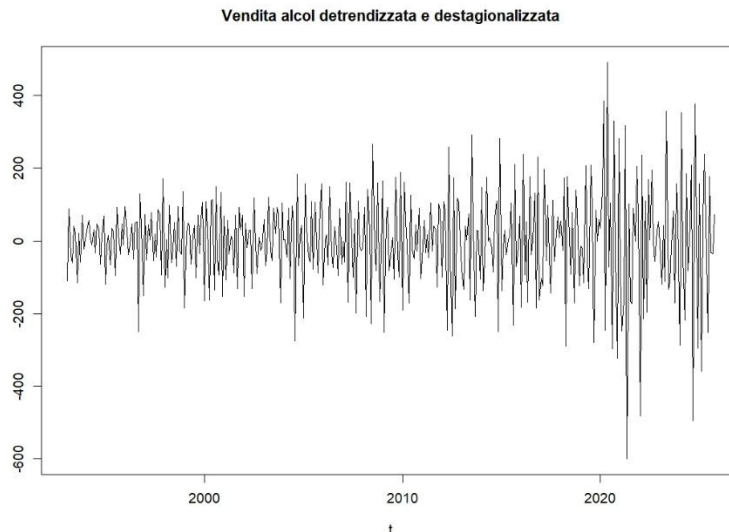
Detrendizzazione e destagionalizzazione

Alla luce dei risultati delle analisi condotte, si decide di detrendizzare e destagionalizzare la serie temporale nel seguente modo:

$$Z_t = \nabla_{12} \nabla X_t$$

```
> dollars_detr <- diff(dollars, lag=1, diff=1)
> dollars_detr_dest <- diff(dollars_detr, lag=12, diff=1)
> plot(dates[-(1:13)], dollars_detr_dest, type="l", ylab="", xlab="t", main="Vendita
alcol detrendizzata e destagionalizzata")
```

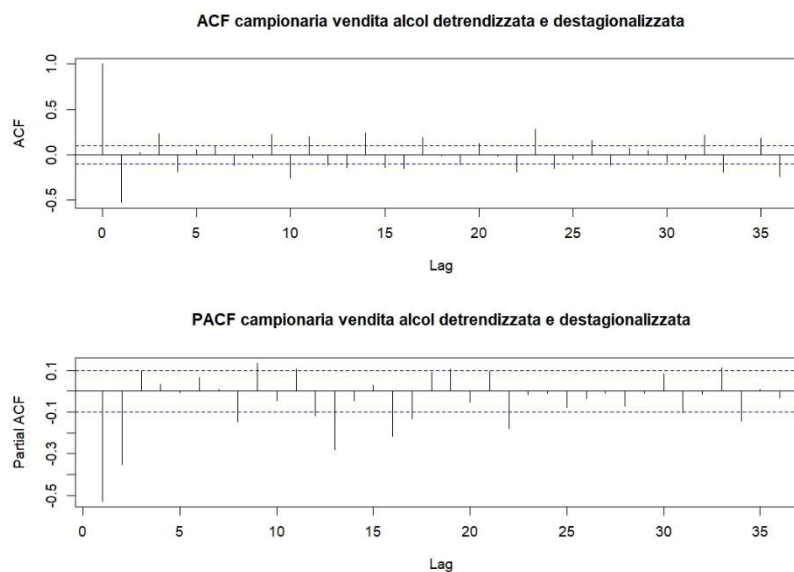
dates[-(1:13)] è il vettore delle date, a cui sono state tolte le prime 13 date. Di seguito è riportato il grafico della serie temporale ottenuta:



Di seguito vengono riportati i correlogrammi:

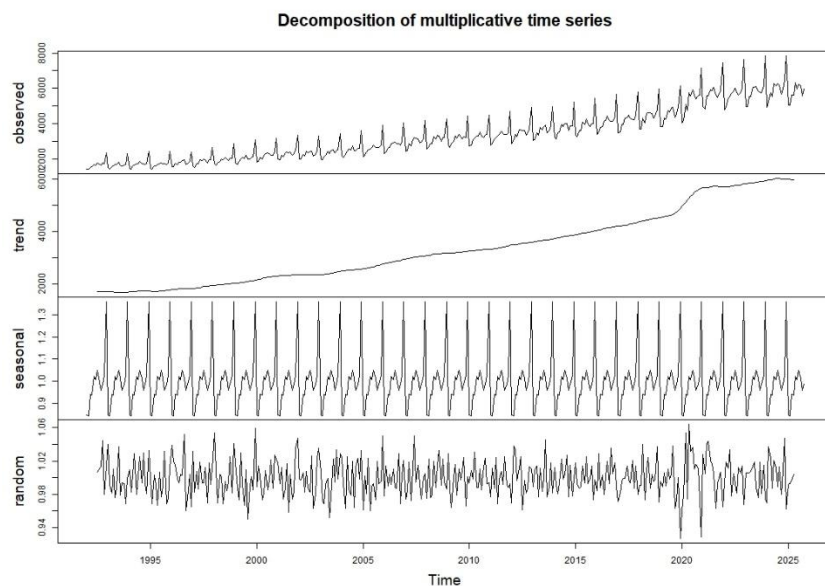
```
> acf(dollars_detr_dest, lag.max=36, main="ACF campionaria vendita alcol
detrendizzata e destagionalizzata")

> pacf(dollars_detr_dest, lag.max=36, main="PACF campionaria vendita alcol
detrendizzata e destagionalizzata")
```



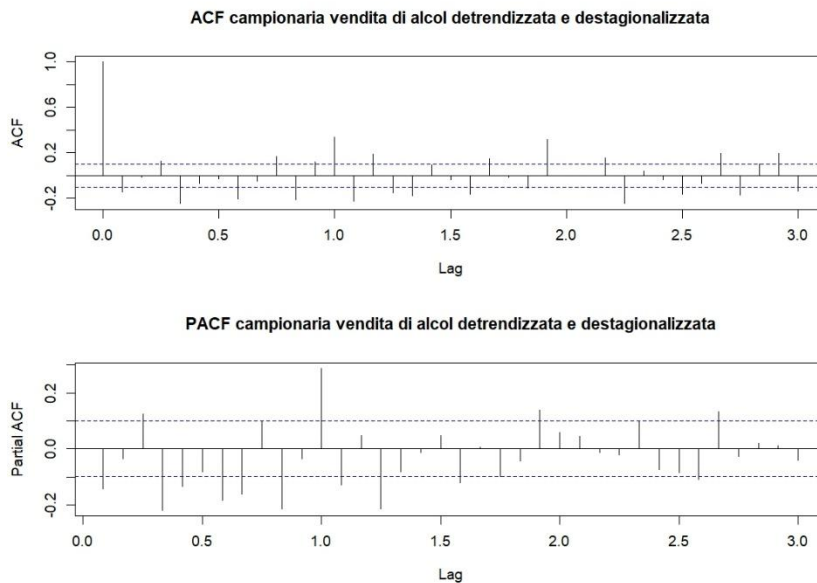
Dai grafici ottenuti è possibile notare che la media del processo è nulla. Tuttavia, la varianza della serie temporale ottenuta non è costante al variare del tempo, ed è molto alta, specialmente a partire dagli anni 2000: ciò suggerisce che il modello sottostante sia moltiplicativo. Inoltre, dai correlogrammi si evince che molte azioni escono dalle bande dell'intervallo di confidenza, motivo per il quale il test di Ljung-Box fallirà quasi sicuramente. Si procede, dunque, a detrendizzare e a destagionalizzare la serie temporale sotto la nuova assunzione che il modello sia moltiplicativo:

```
> dollars_detr_dest <- decompose(dollars_ts, type="multiplicative")
> plot(dollars_detr_dest)
```



Dal grafico della serie “random” si nota che la varianza della serie temporale è pressoché costante. Si procede con l’analisi dei correlogrammi:

```
> serie_pulita <- na.omit(dollars_detr_dest$random)
> par(mfrow=c(2,1))
> acf(serie_pulita, lag.max=36, main="ACF campionaria vendita di alcol detrendizzata
e destagionalizzata")
> pacf(serie_pulita, lag.max=36, main="PACF campionaria vendita di alcol
detrendizzata e destagionalizzata")
```



Purtroppo neanche la decomposizione con modello moltiplicativo ha reso la serie temporale stazionaria. Per confermare i dati riportati dai correlogrammi, si esegue il test di Ljung-Box:

```
> Box.test(serie_pulita, lag=20, type="Ljung-Box")
```

Box-Ljung test

data: serie_pulita

X-squared = 222.48, df = 20, p-value < 2.2e-16

Come previsto, il test rifiuta l'ipotesi che la serie ottenuta sia white noise. Ciò implica che la detrendizzazione e la destagionalizzazione della serie non hanno eliminato alcune componenti del modello dotate di memoria.

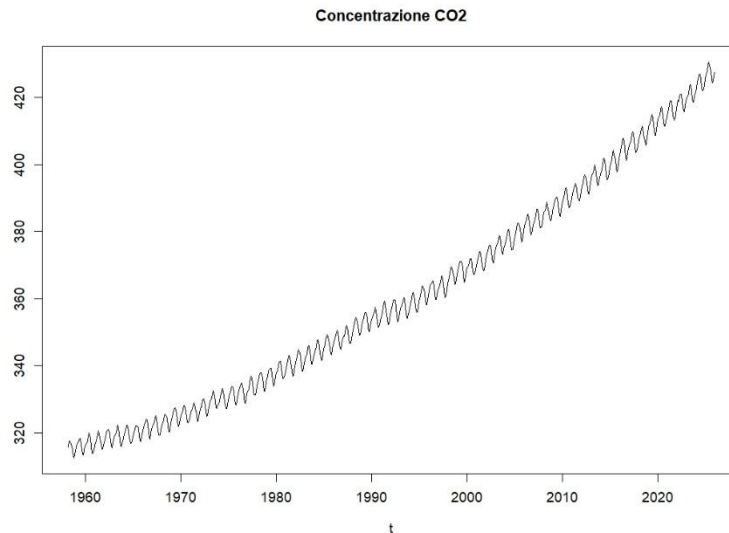
Concentrazione CO2

Il dataset, scaricato dal [sito](#) del Global Monitoring Laboratory, contiene i valori medi della concentrazione di anidride carbonica nell'atmosfera da marzo del 1958 a dicembre del 2025. I dati sono stati raccolti su base mensile. Il dataset contiene diverse voci: di queste, sono state selezionate il mese, l'anno, e il valore medio di anidride carbonica:

```
> co2_data <- read.csv("./co2_mm_mlo.csv", sep=";", comment.char="#")
> co2_values <- co2_data$average
> starting_year <- co2_data$year[1]
> starting_month <- co2_data$month[1]
```

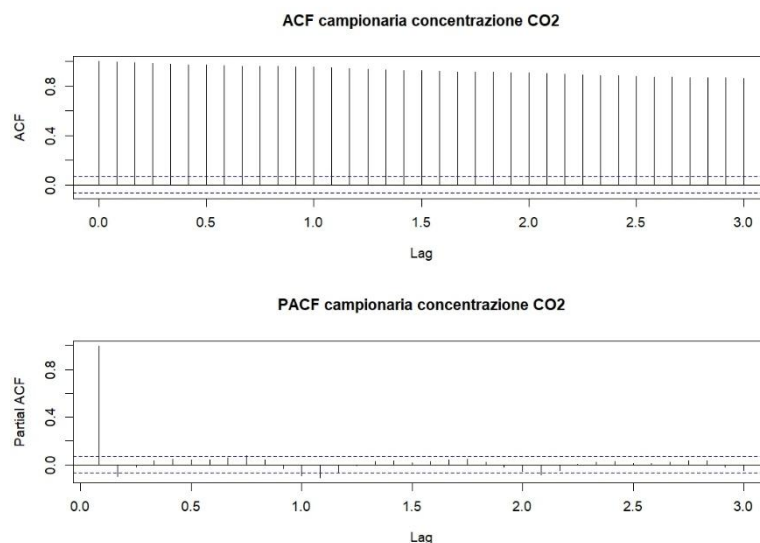


```
> co2_ts <- ts(co2_values, start=c(starting_year, starting_month), frequency=12)
> plot(co2_ts, type="l", ylab="", xlab="t", main="Concentrazione CO2")
```



Dal grafico è possibile notare la presenza di un trend crescente, e di una stagionalità marcata. Inoltre, al variare del tempo la varianza della serie temporale è costante, per si ipotizza che il modello alla base della serie temporale sia di tipo additivo.

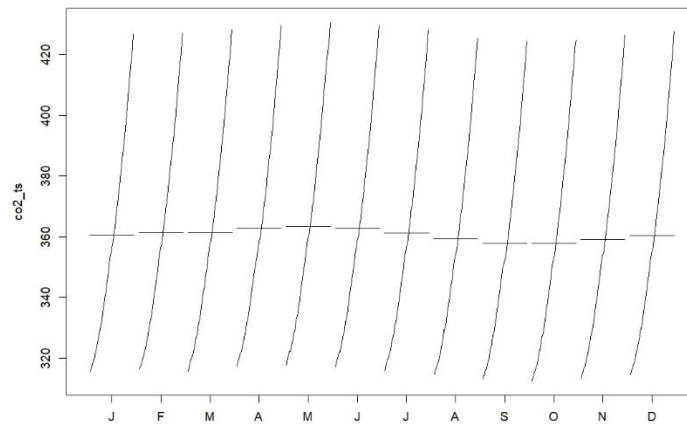
```
> par(mfrow=c(2,1))
> acf(co2_ts, lag.max=36, main="ACF campionaria concentrazione CO2")
> pacf(co2_ts, lag.max=36, main="PACF campionaria concentrazione CO2")
```



La funzione di autocorrelazione evidenzia una decrescita lineare della memoria della serie temporale, mentre dalla funzione di autocorrelazione parziale è possibile notare che il valore della serie temporale in istanti successivi dipende poco dall'istante corrente.

Monthplot

```
> monthplot(co2_ts)
```



Il monthplot conferma il trend crescente e lineare. Inoltre, le medie mensili delle serie storiche hanno valori diversi, seppur non di molto. Per approfondire la stagionalità della serie temporale, si ricorre al seasonal plot.

Seasonal plot

Per tracciare il seasonal plot sono state considerate finestre temporali che partono dal 1960, e sono scandite ogni 10 anni.

```
> y1 <- window(co2_ts, start=c(1960,1), end=c(1960,12))
> y2 <- window(co2_ts, start=c(1970,1), end=c(1970,12))
> y3 <- window(co2_ts, start=c(1980,1), end=c(1980,12))
> y4 <- window(co2_ts, start=c(1990,1), end=c(1990,12))
> y5 <- window(co2_ts, start=c(2000,1), end=c(2000,12))
> y6 <- window(co2_ts, start=c(2010,1), end=c(2010,12))
> y7 <- window(co2_ts, start=c(2020,1), end=c(2020,12))
> range(co2_ts)
```

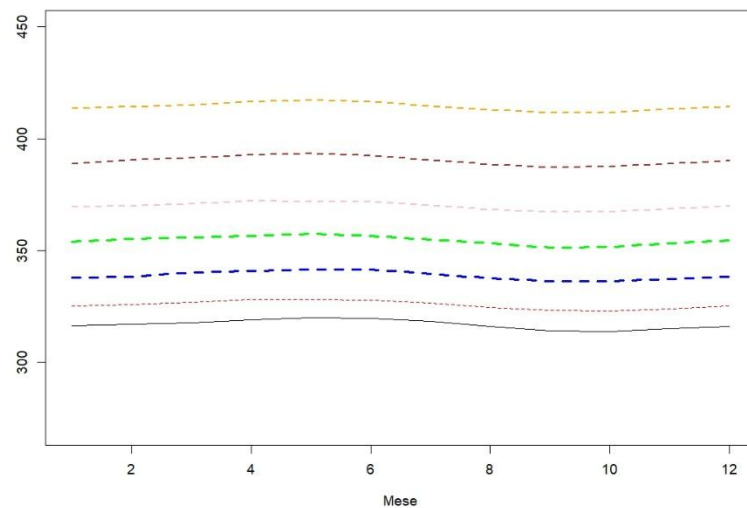
```
[1] 312.42 430.51
```

```
> x <- seq(1, 12, 1)
> matplot(x, y1, type="l", ylim=c(270,450), ylab="", xlab="Mese")
> lines(x, y2, type="l", lty=2, col="red")
```

```

> lines(x, y3, type="l", lty=2, col="blue", lwd=3)
> lines(x, y4, type="l", lty=2, col="green", lwd=3)
> lines(x, y5, type="l", lty=2, col="pink", lwd=2)
> lines(x, y6, type="l", lty=2, col="brown", lwd=2)
> lines(x, y7, type="l", lty=2, col="orange", lwd=2)

```



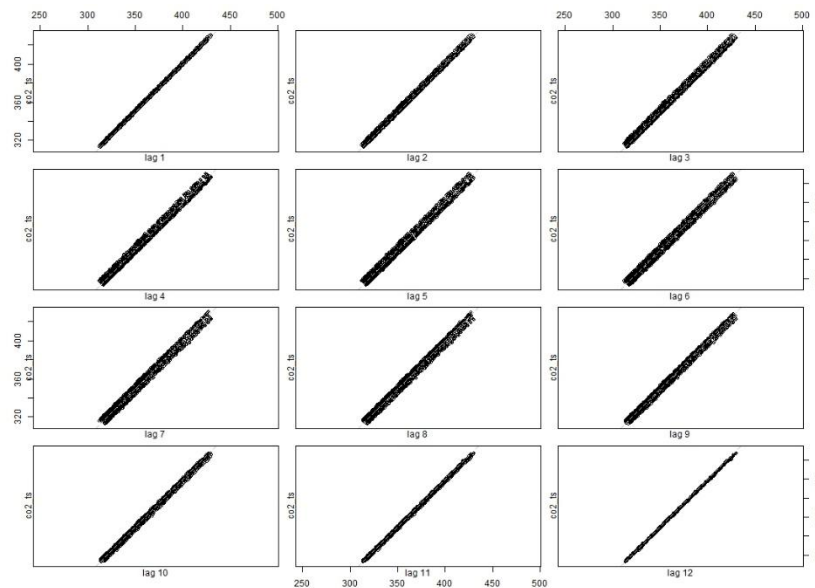
Dal seasonal plot è possibile notare che i grafici hanno incrementi che aumentano sempre di più con l'avanzare degli anni. Inoltre, i grafici hanno una forma molto simile tra di loro. Quindi, per comprendere il tipo di stagionalità, si ricorre al lag plot.

Lag plot

```

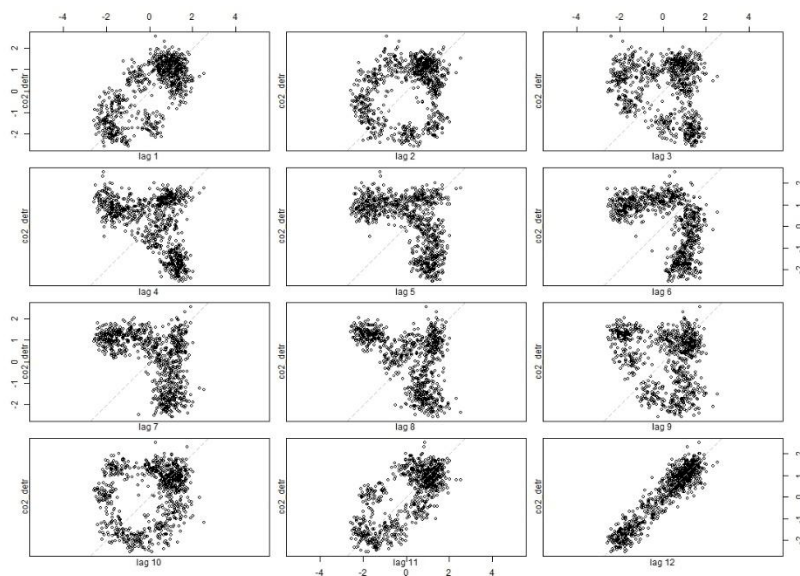
> lag.plot(co2_ts, set.lags=1:12, type="p", do.lines=FALSE)

```



Anche il lag plot conferma la presenza del trend lineare. Inoltre, in ciascun lag i punti si dispongono lungo una retta, più o meno spessa, a indicare che il trend è fortemente presente all'interno dei dati. Osservando il grafico corrispondente al lag 12, si nota che i punti sono disposti lungo una retta molto sottile, ed è il segno della stagionalità. Infatti, detrendizzando la serie temporale, e ridisegnando il lag plot, si ottiene:

```
> co2_detr <- diff(co2_ts, lag=1, differences=1)
> lag.plot(co2_detr, set.lags=1:12, type="p", do.lines=FALSE)
```



I lag che vanno da 1 a 11 ora sono più dispersi, mentre nel lag 12 rimane la concentrazione dei punti lungo la retta. Alla luce di ciò, è possibile concludere la presenza di un trend lineare, e di una stagionalità annuale.

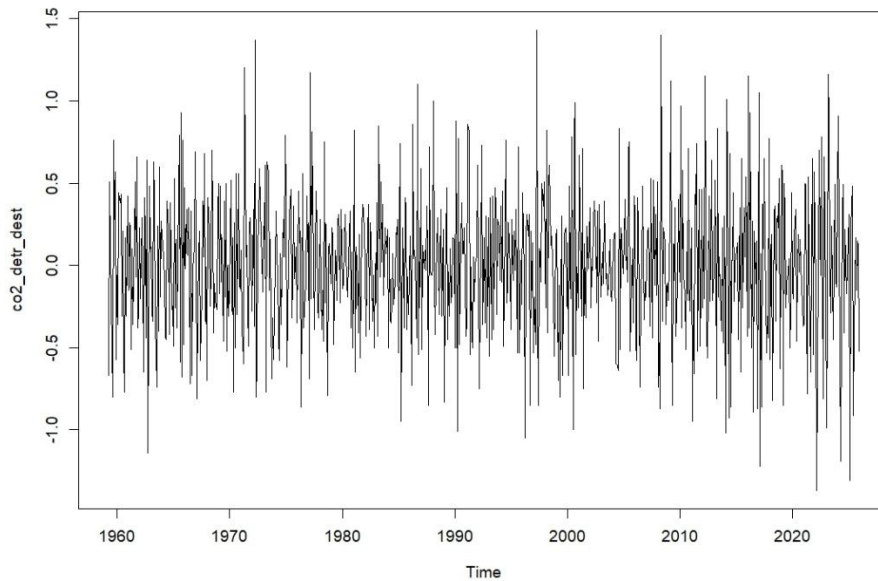
```
> co2_detr <- diff(co2_ts, lag=1, differences=1)
```

```
> co2_detr_dest <- diff(co2_detr, lag=12, differences=1)
> plot(co2_detr_dest)
```

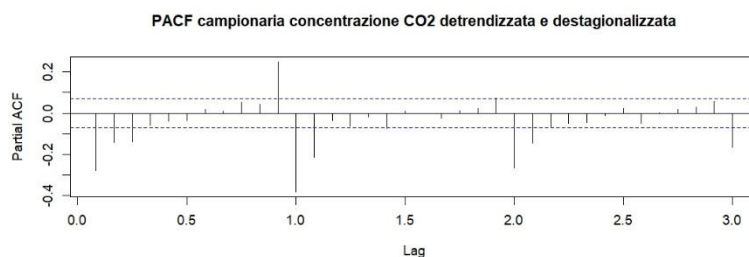
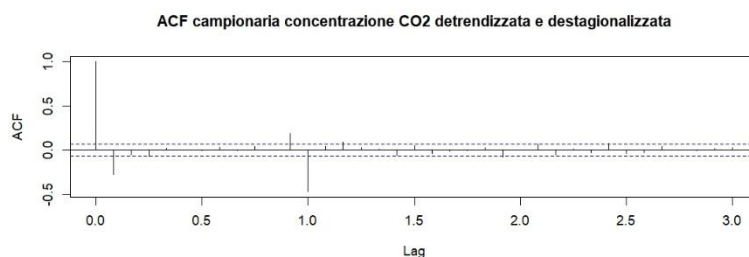
Detrendizzazione e destagionalizzazione

La serie temporale è detrendizzata e destagionalizzata nel seguente modo:

$$Z_t = \nabla_{12} \nabla X_t$$



```
> par(mfrow=c(2,1))
> acf(co2_detr_dest, lag.max=36, main="ACF campionaria concentrazione CO2
detrendizzata e destagionalizzata")
> pacf(co2_detr_dest, lag.max=36, main="PACF campionaria concentrazione CO2
detrendizzata e destagionalizzata")
```



Dalla funzione di autocorrelazione parziale è possibile notare che sono ancora presenti delle correlazioni all'interno della serie temporale detrendizzata e destagionalizzata.

Eseguendo il test di Ljung-Box, si ottiene:

```
> Box.test(co2_detr_dest, lag=20, type="Ljung-Box")
```

Box-Ljung test

data: co2_detr_dest

X-squared = 299.82, df = 20, p-value < 2.2e-16

Dunque, anche in questo caso viene rifiutata l'ipotesi che la serie temporale detrendizzata e destagionalizzata sia un white noise. Il motivo potrebbe essere dovuto a fenomeni anomali, come incendi o eruzioni vulcaniche, che hanno aumentato temporaneamente la concentrazione di anidride carbonica.