

# Analisi serie temporali tramite software R

<b>Studi Preliminari</b>	<b>1</b>
Simulazione Processo White Noise	1
Simulazione Processo AR(1) & MA(1)	3
Trend & Stagionalità	6
Studio preliminare stagionalità	13
<b>Matrimoni secondo il rito - Anni 1948-2014</b>	<b>17</b>
<b>Opere librerie pubblicate per materia trattata(1926-2014)</b>	<b>24</b>
<b>Stima delle emissioni inquinanti in atmosfera dal 1980-2010</b>	<b>49</b>

## Studi Preliminari

In questo testo verranno mostrate le principali operazioni che tramite il software R permettono di analizzare serie storiche estrapolando eventuali tendenze e componenti presenti, al fine di poter effettuare eventuali predizioni sul comportamento di un determinato fattore descritto dalla serie.

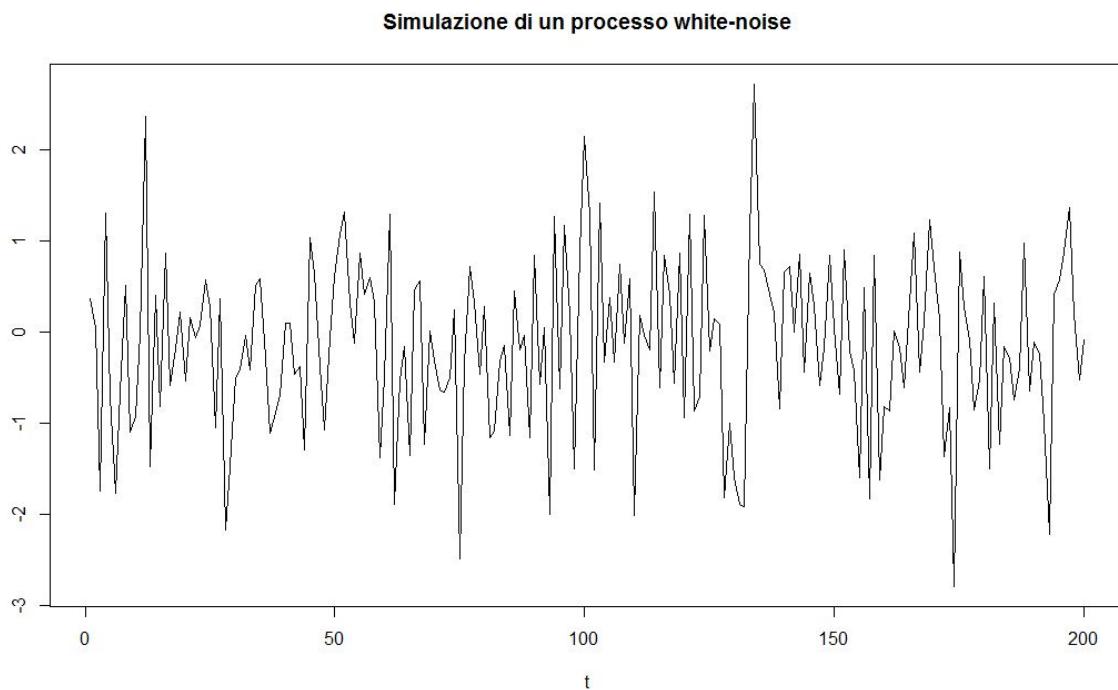
In ordine verranno mostrati prima le analisi relative ad alcuni dataset nativi di R in modo da poter riscontrare eventuali differenze negli output, successivamente si analizzeranno serie storiche fornite dall'ISTAT.

## Simulazione Processo White Noise

Il comando **arima.sim()** permette di effettuare la simulazione di un processo ARIMA. L'argomento passato alla funzione **order()** permette di indicare l'ordine del processo **ARIMA(p,d,q)**. Per simulare un processo white-noise(WN) bisogna impostare a 0 i parametri dell'argomento ( $p=q=0$ ).

Tramite **n** invece si indica ad R il numero di realizzazioni desiderate.

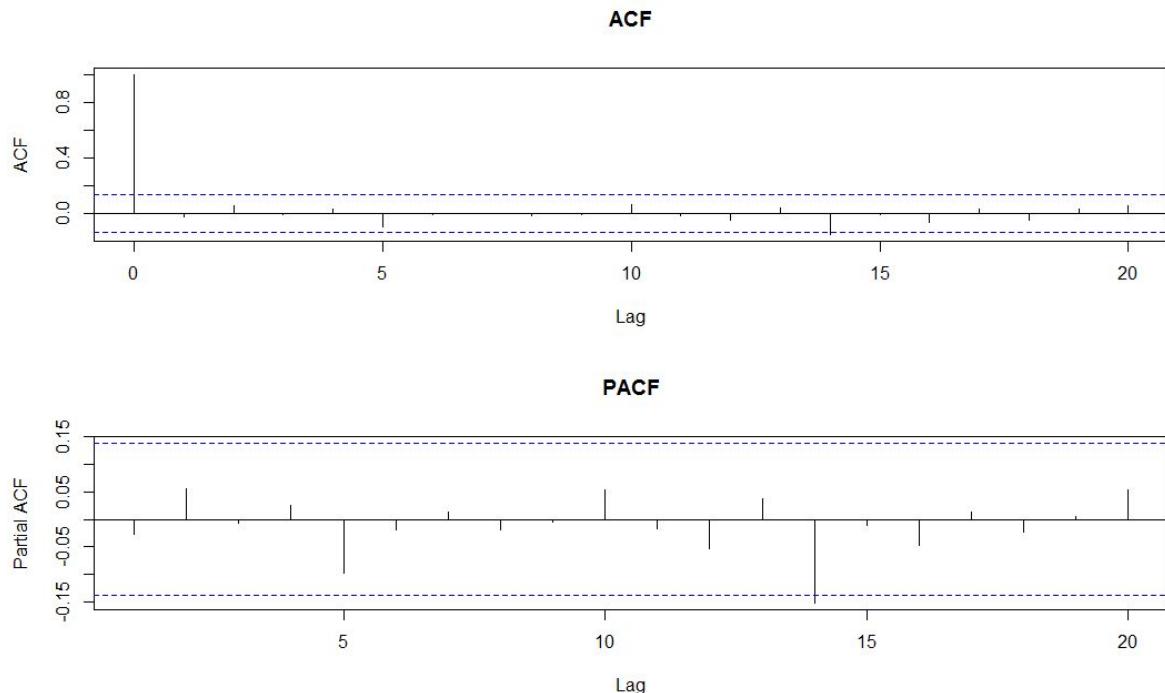
```
> eps<-arima.sim(list(order=c(0,0,0)), n=200)
> plot(eps, ylab= " ", xlab="t", main="Simulazione di un processo white-noise")
```



E' possibile osservare come il processo vari attorno allo 0 con oscillazioni di ampiezza più o meno costante; infatti risulta  $E(\varepsilon_t) = 0$  e  $\text{Var}(\varepsilon_t) = \sigma^2$ .

Le funzioni di autocorrelazione globale e autocorrelazione parziale, relative al campione, sono ottenute tramite i successivi comandi in cui l'argomento **lag.max** indica i ritardi in corrispondenza dei quali si desidera calcolare l'autocorrelazione; dunque nei grafici si riportano i valori di  $\rho(1), \rho(2), \dots, \rho(20)$ .

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



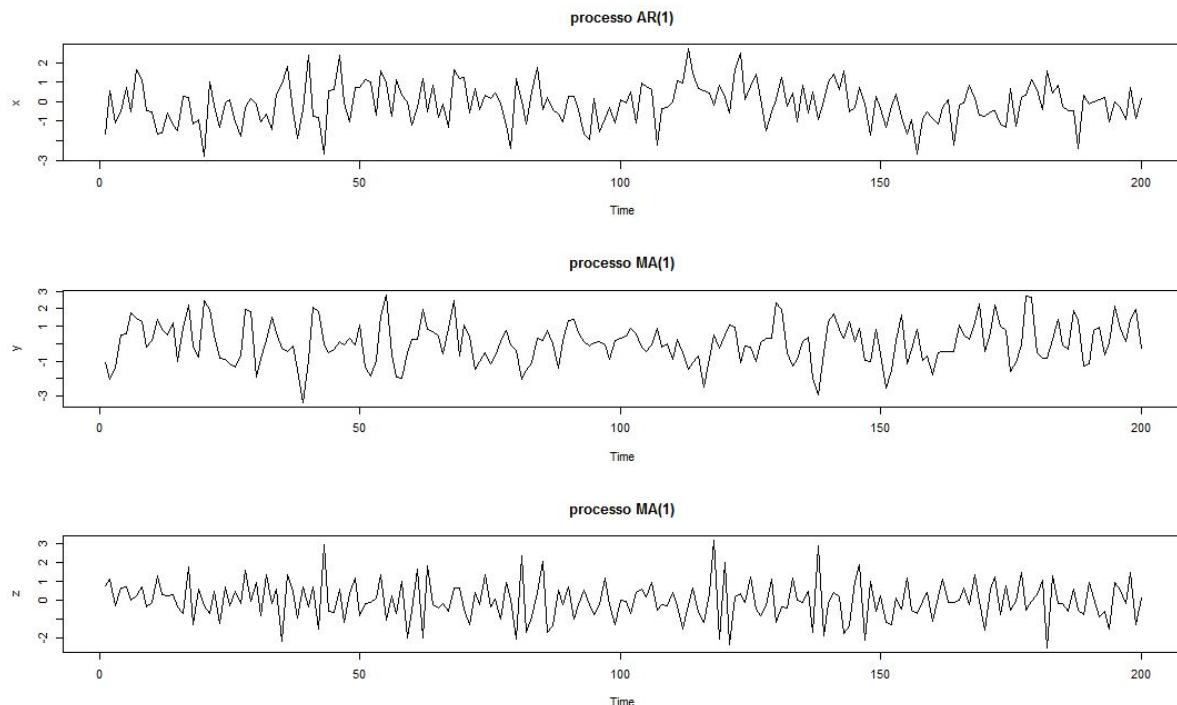
## Simulazione Processo AR(1) & MA(1)

Si passa ora a simulare i processi AR(1) e MA(1):

$$\begin{aligned} X_t &= 0.2 X_{t-1} + \varepsilon_t \\ Y_t &= \varepsilon_t + 0.5 \varepsilon_{t-1} \\ Z_t &= \varepsilon_t - 0.5 \varepsilon_{t-1} \end{aligned}$$

tramite i seguenti comandi:

```
> x<-arima.sim(list(order=c(1,0,0), ar=0.2), n=200)
> y<-arima.sim(list(order=c(0,0,1), ma=0.5), n=200)
> z<-arima.sim(list(order=c(0,0,1), ma=-0.5), n=200)
> par(mfrow=c(3,1))
> plot(x, type="l", main="processo AR(1)")
> plot(y, type="l", main="processo MA(1)")
> plot(z, type="l", main="processo MA(1)")
```

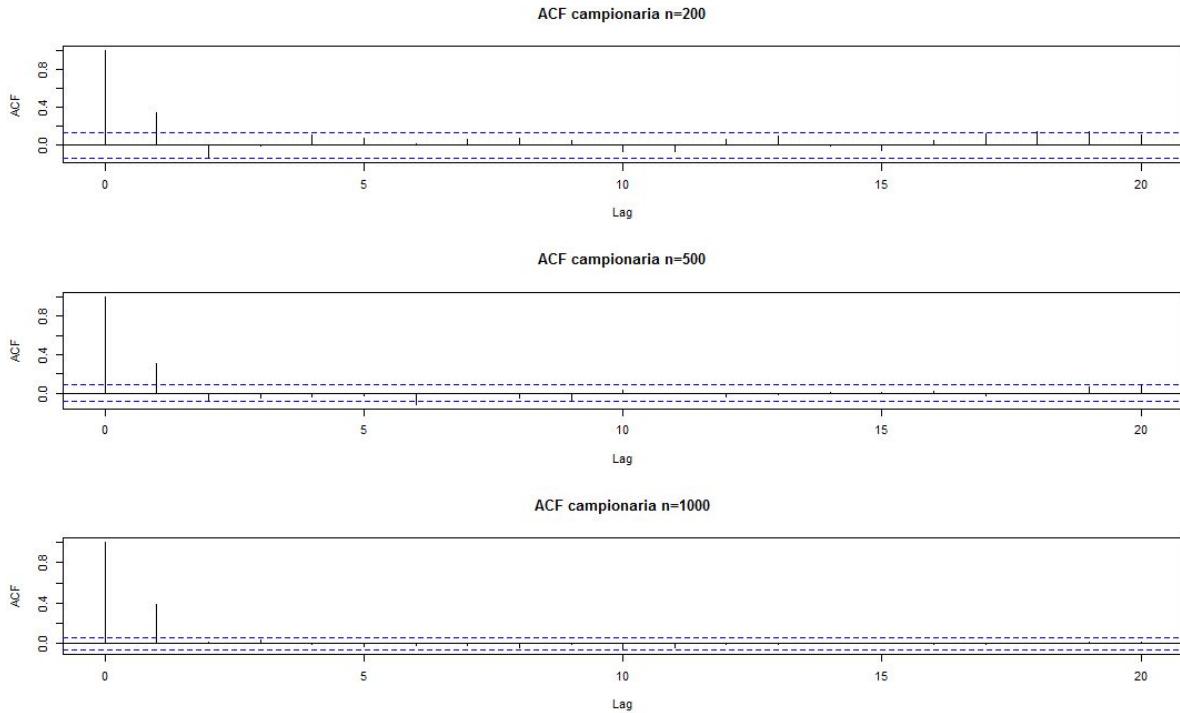


Il processo  $Z_t$ , sebbene differisca da  $Y_t$  solo per il segno del coefficiente della componente *white noise*, compie rispetto a quest'ultimo oscillazioni con una frequenza maggiore; ciò è dovuto alla autocorrelazione negativa che caratterizza la funzione di autocorrelazione della  $Z_t$ .

Di seguito si procede a simulare dei processi MA(1), come  $Y_t$ , per vedere come al crescere del campione la funzione di autocorrelazione si avvicina a quella teorica, che risulta essere:

$$\rho(h) = \begin{cases} 1 & h = 0 \\ 0.4 & h = 1 \\ 0 & h = 2,3,\dots \end{cases}$$

```
> y1<-arima.sim(list(order=c(0,0,1), ma=0.5), n=200)
> y2<-arima.sim(list(order=c(0,0,1), ma=0.5), n=500)
> y3<-arima.sim(list(order=c(0,0,1), ma=0.5), n=1000)
> par(mfrow=c(3,1))
> acf(y1, lag.max =20, main="ACF campionaria n=200")
> acf(y2, lag.max =20, main="ACF campionaria n=500")
> acf(y3, lag.max =20, main="ACF campionaria n=1000")
```



La funzione di autocorrelazione può essere determinata anche in modo esatto, tramite il comando **ARMAacf()**, in quanto l'autocorrelazione è nota e ha forma esplicita per un qualsiasi processo MA(q).

$$X_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q}$$

Ad esempio per il processo MA(3) seguente, l'immagine ci mostra l'output per il calcolo dell'autocorrelazione

$$X_t = \varepsilon_t + 0.2 \varepsilon_{t-1} - 0.15 \varepsilon_{t-2} + 0.2 \varepsilon_{t-3}$$

```
> r<-ARMAacf(0,c(0.2,-0.15,0.2),lag.max = 15)
> r
      0         1         2         3         4         5         6         7         8
1.000000000 0.12698413 -0.09977324 0.18140590 0.00000000 0.00000000 0.00000000
      9        10        11        12        13        14        15
0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
```

Risulta più utile calcolare l'autocorrelazione nel caso questa non sia espressa da una formula chiusa, ossia ad esempio nel caso di un **ARMA(p,q)** con  $p \geq 1$  e  $q \geq 1$ .

$$X_t = 0.8 X_{t-1} + \varepsilon_t - 0.45 \varepsilon_{t-1}$$

Dunque l'output dell'esempio diventa:

```

> r<-ARMAacf(ar=0.8, ma=-0.45, lag.max = 15)
> r
      0      1      2      3      4      5      6      7      8
1.00000000 0.46424870 0.37139896 0.29711917 0.23769534 0.19015627 0.15212502 0.12170001 0.09736001
      9      10     11     12     13     14     15
0.07788801 0.06231041 0.04984833 0.03987866 0.03190293 0.02552234 0.02041787

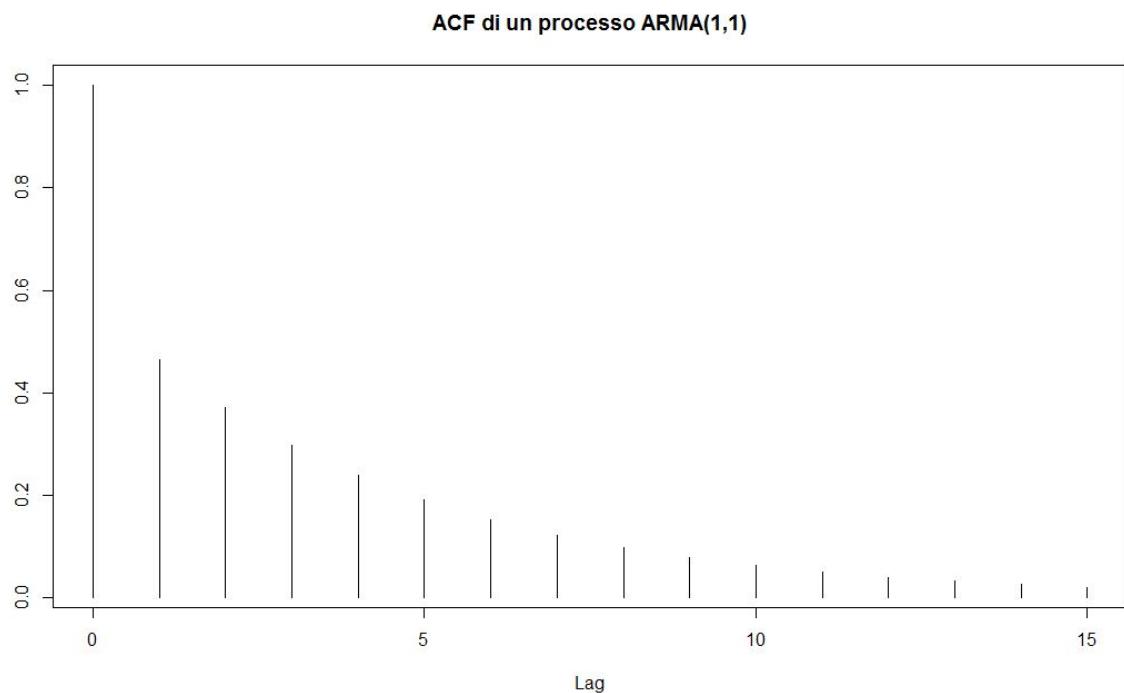
```

Segue il grafico del relativo andamento:

```

> h<-seq(0,15,1)
> matplot(h,r, ylab="", xlab="Lag", type="h", main="ACF di un processo ARMA(1,1)")

```



Infine mediante un parametro opportuno è possibile calcolare e poi tracciare il grafico della funzione di autocorrelazione parziale.

```

> r<-ARMAacf(ar=0.8,ma=-0.45,lag.max = 15, pacf = TRUE)
> r
[1] 4.642487e-01 1.986965e-01 8.853677e-02 3.976260e-02 1.788599e-02 8.048043e-03 3.621560e-03 1.629696e-03
[9] 7.333629e-04 3.300133e-04 1.485060e-04 6.682768e-05 3.007246e-05 1.353261e-05 6.089673e-06

```

## Trend & Stagionalità

In una serie storica, la presenza di trend o stagionalità spesso compromette la stazionarietà, per cui si procede a “detrendizzare” la serie al fine di individuare quale modello vi si adatti meglio. Dunque è possibile usare filtri lineari, ossia medie mobili di ordine  $m$ , in cui la stima del trend in  $t$  è data da:

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

Invece, nel caso in cui i dati presentino stagionalità, la stima si modifica nella seguente espressione (in cui per dati che siano trimestrali si usa  $m=2$  , mentre per dati mensili si usa  $m=6$  )

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

Tali filtri lineari sono messi a disposizione da R attraverso la funzione filter() in cui occorre specificare, innanzitutto, il vettore dei coefficienti con cui vengono pesate le osservazioni nel filtro. Ad esempio, per una media mobile semplice ponderata di ordine  $m$  è possibile utilizzare un vettore di coefficienti fatto come segue:

$$\alpha = \left( \frac{1}{2m+1}, \dots, \frac{1}{2m+1} \right)$$

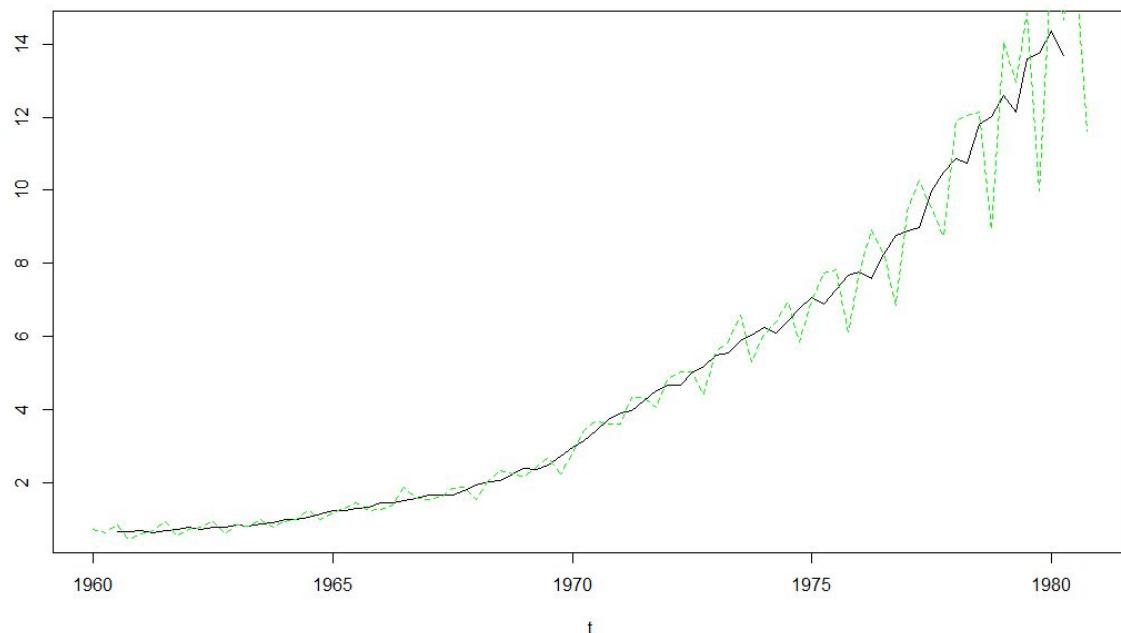
Il vettore dei coefficienti viene memorizzato da R, mediante il comando rep a cui vengono passati il valore e il numero di elementi (la somma deve essere pari ad 1).

Segue un'applicazione di quanto detto sulla detrendizzazione, fatta sul dataset JohnsonJohnson presente in R (guadagni trimestrali dell'azienda dal 1960 al 1980). Si ritiene che tale analisi possa essere più significativa rispetto a quella effettuabile sul PIL, i cui risultati sono già riportati nelle dispense. Tali dati risultano essere, come quelli sul PIL, trimestrali.

```
> data(JohnsonJohnson)
> alpha<-rep(1/5,5)
> ma <- filter(JohnsonJohnson, gamma, method="convolution", sides=2, circular=FALSE)
> plot(ma, ylab=" ", xlab="t", main="Media mobile di ordine 2 per la serie JohnsonJohnson")
> lines(JohnsonJohnson, col="green", lty=2)
```

Nei comandi esposti viene fatta una media mobile semplice di ordine 2 sui dati considerati: viene importato il dataset, vengono memorizzati i pesi (strutturati nel modo precedentemente esposto) e infine viene applicato il filtro vero e proprio tramite la funzione filter(). A questa funzione vengono passati, come argomenti, la serie storica, il vettore dei pesi, il metodo - che per una media mobile risulta essere 'convolution'-, il parametro sides - che indica che le osservazioni utilizzate dal filtro sono le  $m$  precedenti e le  $m$  successive a  $t$ , compresa quella in  $t$  stessa. Il parametro circular , infine, serve per porre NA (Not Available) in corrispondenza delle osservazioni perse all'inizio e alla fine (vengono perse  $2*m$  osservazioni, come riportato nella figura seguente).

Media mobile di ordine 2 per la serie JohnsonJohnson



	Qtr1	Qtr2	Qtr3	Qtr4
1960	NA	NA	0.648	0.644
1961	0.702	0.642	0.698	0.730
1962	0.776	0.712	0.768	0.784
1963	0.830	0.800	0.864	0.898
1964	0.986	0.986	1.064	1.140
1965	1.230	1.232	1.284	1.328
1966	1.440	1.462	1.518	1.584
1967	1.674	1.674	1.668	1.776
1968	1.926	2.010	2.070	2.250
1969	2.376	2.358	2.466	2.718
1970	2.970	3.150	3.420	3.726
1971	3.906	3.978	4.230	4.518
1972	4.662	4.680	4.986	5.184
1973	5.490	5.544	5.868	6.030
1974	6.246	6.102	6.426	6.768
1975	7.056	6.894	7.272	7.668
1976	7.776	7.578	8.262	8.766
1977	8.892	8.982	9.990	10.494
1978	10.872	10.746	11.808	12.024
1979	12.582	12.150	13.608	13.734
1980	14.346	13.698	NA	NA

In realtà, essendo i dati trimestrali, occorrerebbe usare più propriamente un media mobile centrata di ordine 2, come la seguente:

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

Per cui il vettore dei coefficienti diventa :  $\alpha=(1/8, 1/4, 1/4, 1/4, 1/8)$

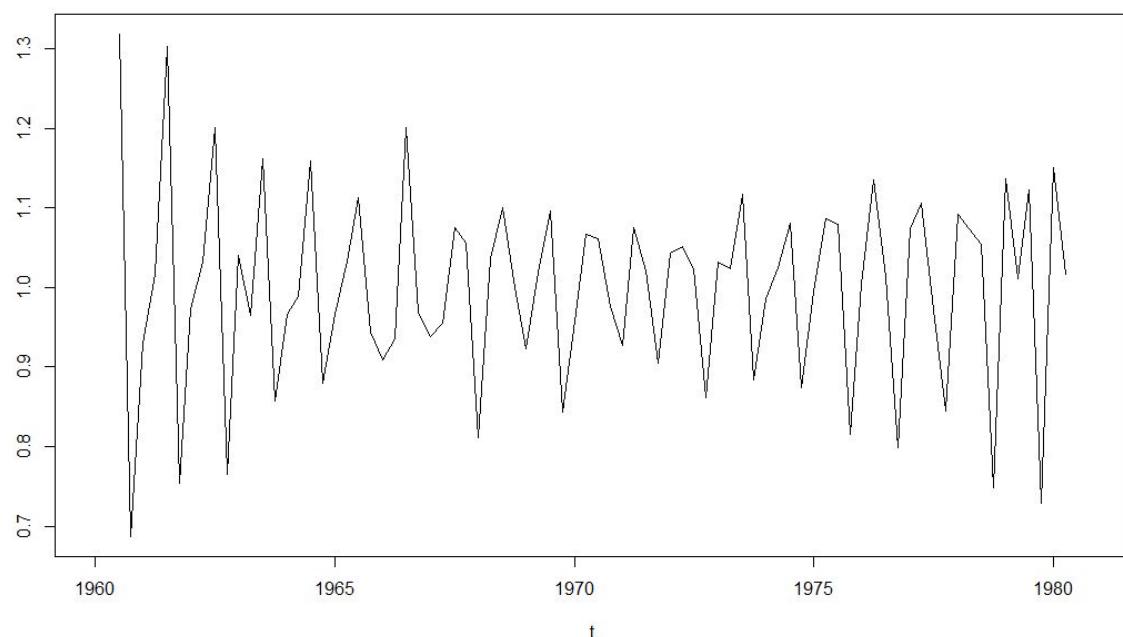
```

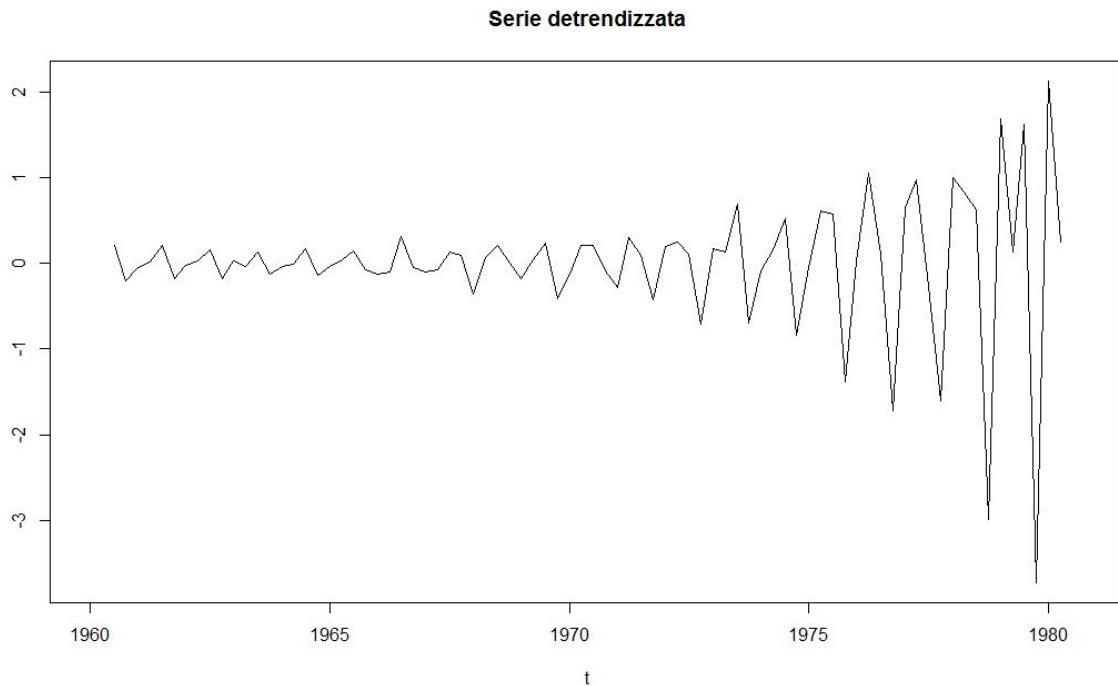
> alpha <- c(1/8, 1/4, 1/4, 1/4, 1/8)
> mac <- filter(JohnsonJohnson, alpha, method="convolution", sides=2, circular=FALSE)
> detrendx <- JohnsonJohnson-mac
> plot(detrendx,ylab=" ",xlab="t",main="Serie detrendizzata")
> detrendxx <- JohnsonJohnson/mac
> plot(detrendxx,ylab=" ",xlab="t",main="Serie detrendizzata modello moltiplicativo")

```

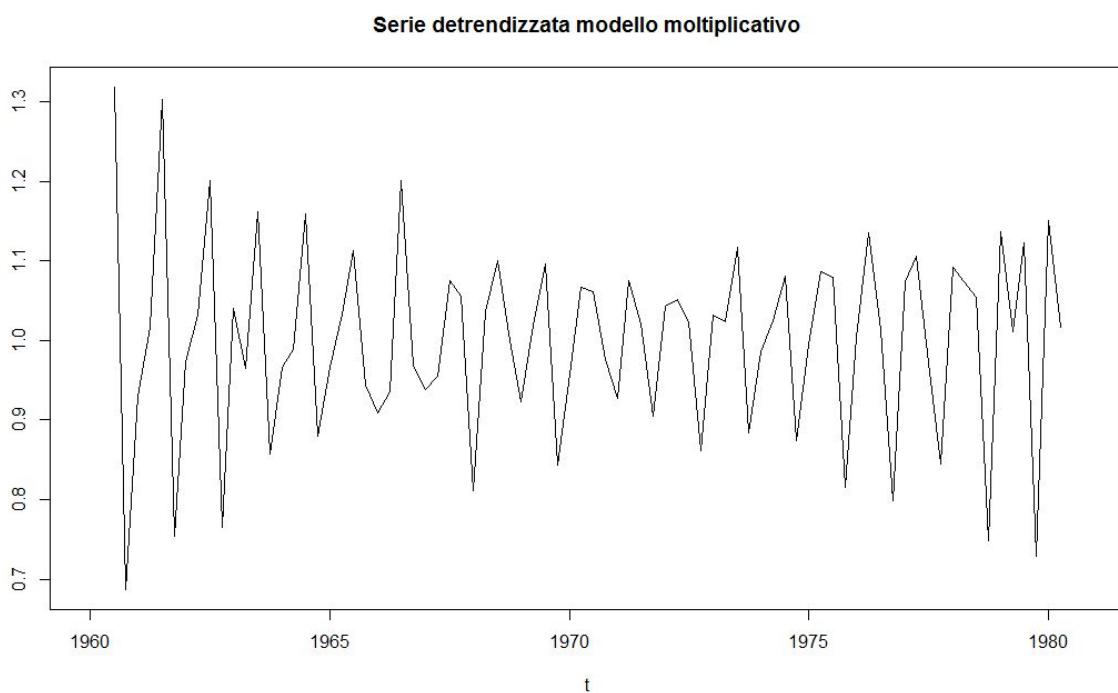
Come si evince dalla figura sovraesposta, i pesi sono cambiati; il tipo di filtro, invece, è rimasto invariato. Si ottiene la serie 'detrendizzata'  $X_t - T_t$  come differenza della serie originale e del trend ricavato precedentemente attraverso la media mobile centrata di ordine 2. Analogamente è possibile effettuare la detrendizzazione considerando come trend quello ricavato con la media mobile semplice ( $ma$ ). Di seguito si effettua un paragone tra le due serie detrendizzate, che risultano tuttavia molto simili.

**Serie detrendizzata modello moltiplicativo**



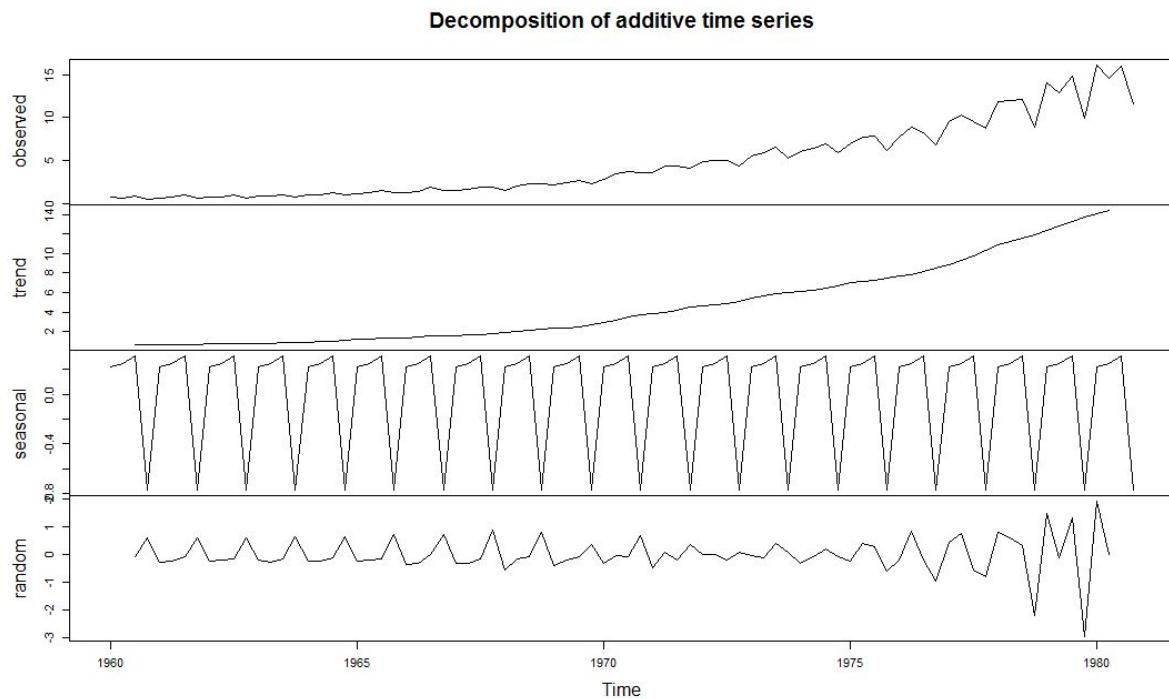


Nel fare questo è stato supposto che il tipo di modello fosse additivo; se il modello fosse moltiplicativo occorrerebbe detrendizzare tramite  $Xt/Tt$ . Nella figura seguente il trend utilizzato è quello stimato tramite media mobile centrata.



E' possibile scomporre una serie storica nelle sue componenti (trend, stagionalità e errore) utilizzando la funzione **decompose()** e specificando il tipo di modello da utilizzare (moltiplicativo o additivo, in base a cui cambia l'espressione dell'errore).

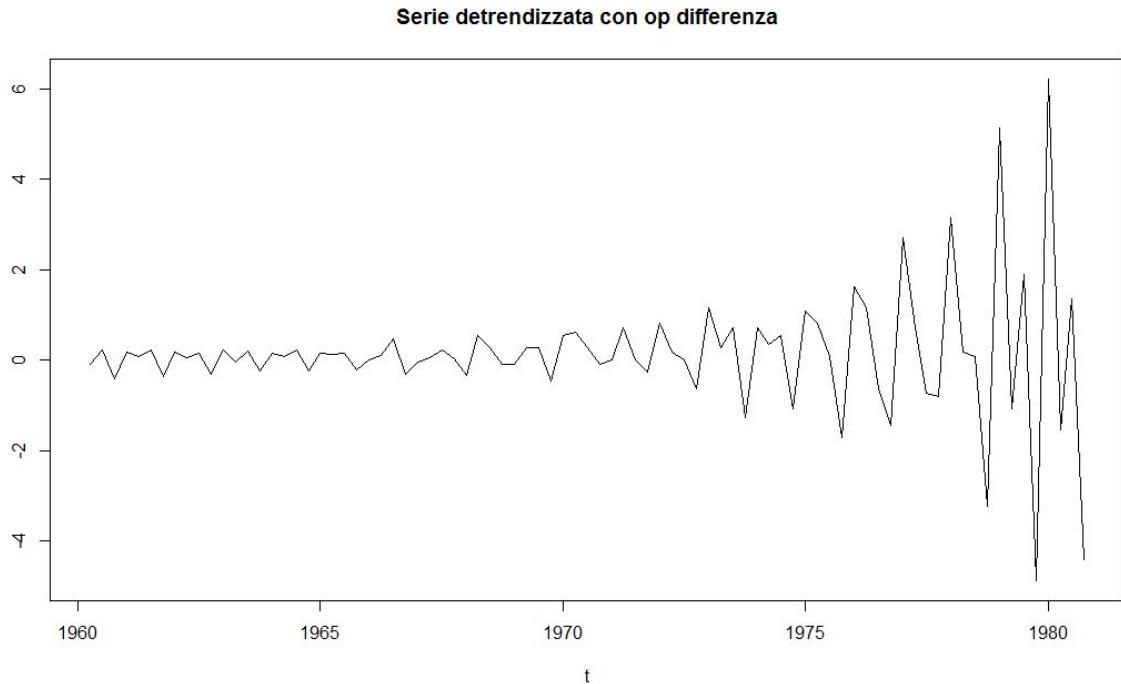
```
> decJ<-decompose(JohnsonJohnson, type="additive")
> plot(decJ)
```



Nell'oggetto **decJ** , R memorizza le diverse componenti: la componente stagionale (seasonal), che per dati trimestrali è composta da 4 stime che rimangono costanti per tutto il periodo di osservazione, la componente di trend (trend) stimata con una media mobile di ordine 2 e, infine, la componente di errore (random) stimata, in accordo con un modello additivo, come differenza tra la serie originale (observed) e le due componenti precedenti.

La detrendizzazione di una serie storica può essere effettuata anche mediante gli operatori differenza: in R è possibile utilizzare tali operatori tramite il comando **diff()**. In tal modo è possibile, oltre che detrendizzare la serie, anche destagionalizzarla, passando opportuni parametri alla funzione **diff()**. Se si suppone che il trend sia lineare, è possibile detrendizzare con il comando seguente, ottenendo un andamento molto simile a quelli già visti.

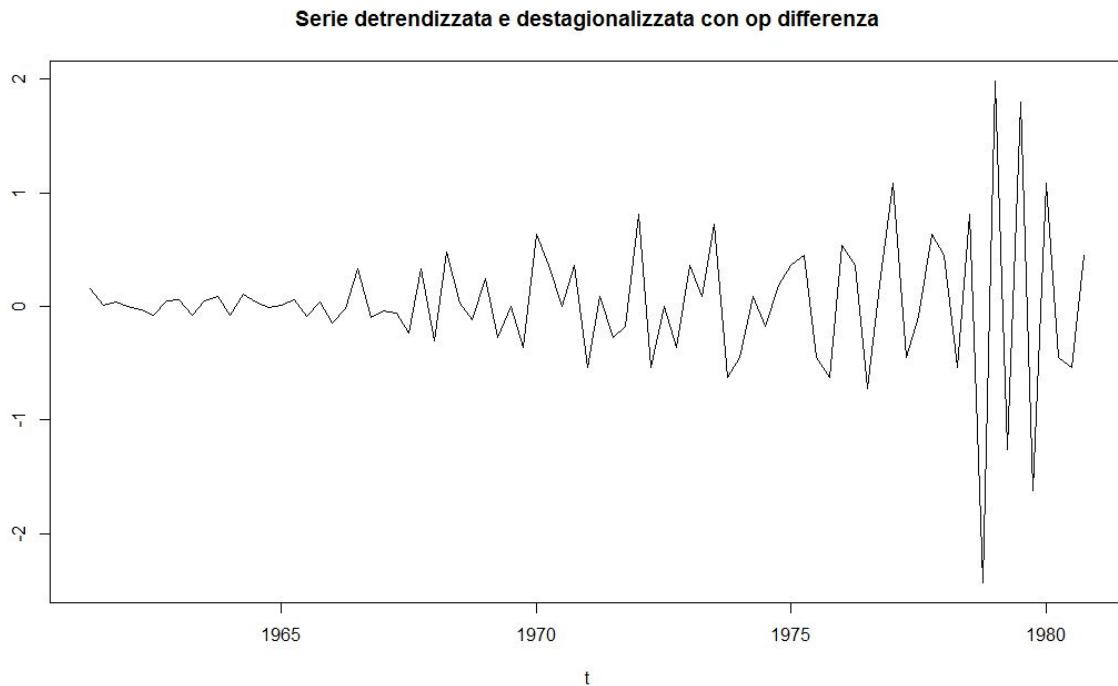
```
> dxt <- diff(JohnsonJohnson, lag=1, differences=1)
> plot(dxt, ylab=" ", xlab="t", main="Serie detrendizzata con op differenza")
```



Tra gli argomenti passati a `diff()`, oltre alla serie originale, **lag** indica la potenza da applicare all'operatore ritardo, mentre `differences` coincide con l'ordine (nel nostro caso risulta pari ad 1, avendo supposto la serie lineare).

Per destagionalizzare, oltre che detrendizzare, occorre cambiare il valore passato del parametro `lag`, che per dati trimestrali deve essere posto pari a 4, mentre per dati mensili deve essere posto pari a 12. Infine, differenze di ordine più elevato si ottengono cambiando il valore del parametro `differences` (e.g per differenziare 2 volte si pone tale parametro pari a 2)

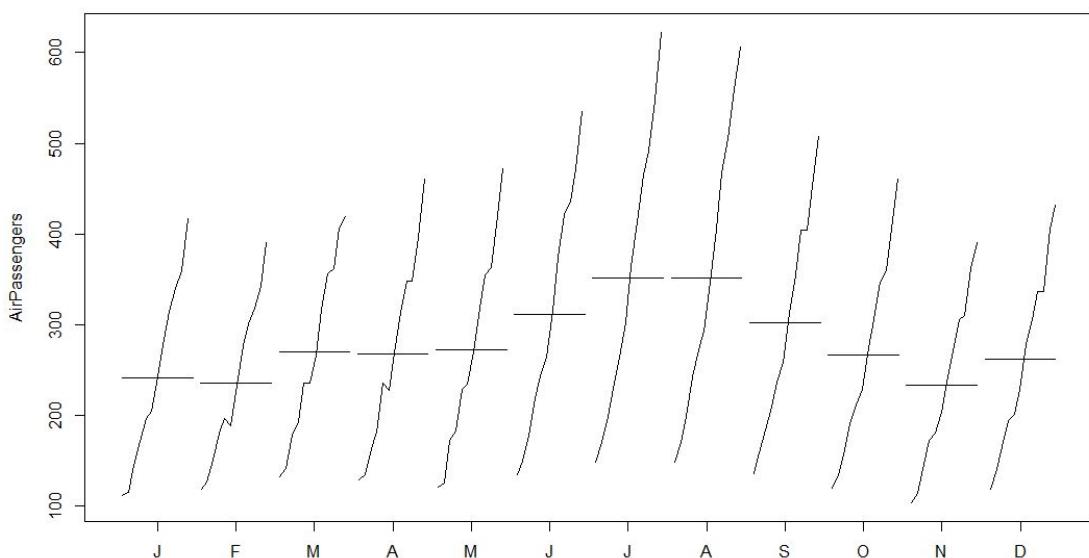
```
> ddxt <- diff(dxt, lag=4, differences=1)
> plot(ddxt, ylab=" ", xlab="t", main="Serie detrendizzata e destagionalizzata con op differenza")
```



## Studio preliminare stagionalità

La stagionalità può essere preliminarmente studiata tramite la funzione **monthplot()**, che su dati mensili genera un grafico per ogni mese dell'anno, ciascuno dei quali rappresenta l'andamento della serie in quel mese. Si applica questa funzione al dataset *AirPassenger*, già presente in R.

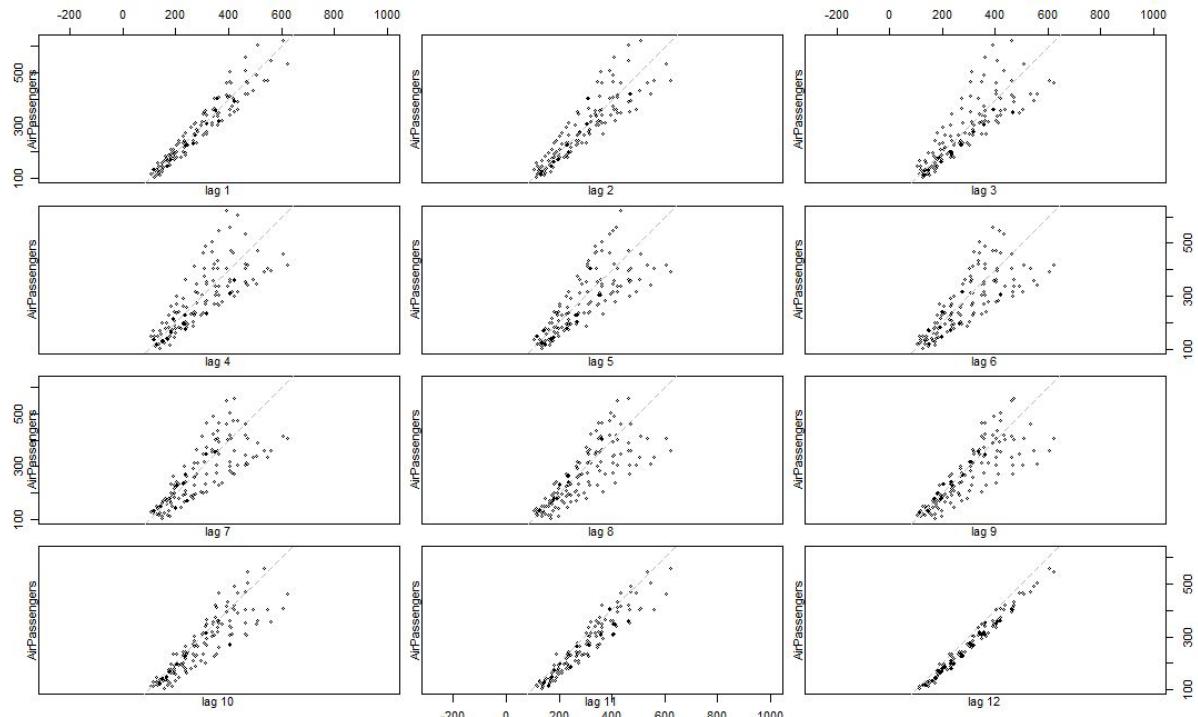
```
> data("AirPassengers")
> monthplot(AirPassengers)
```



Dal grafico precedente, che indica la media di ogni sottoserie individuata per ciascun mese, si vede che nei mesi di Luglio e Agosto vi è un incremento sensibile del numero di passeggeri. In generale lo shift delle diverse linee, l'una rispetto all'altra, rappresenta il trend della serie complessiva.

Per individuare trend e/o stagionalità è possibile usare anche un grafico lagplot che R mette a disposizione mediante l'omonima funzione (**lagplot()**). Questo grafico è un grafico di autodispersione costituito da coppie di punti  $\{(X_t, X_{t+k}): t=1, 2, \dots, n-k\}$ .

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

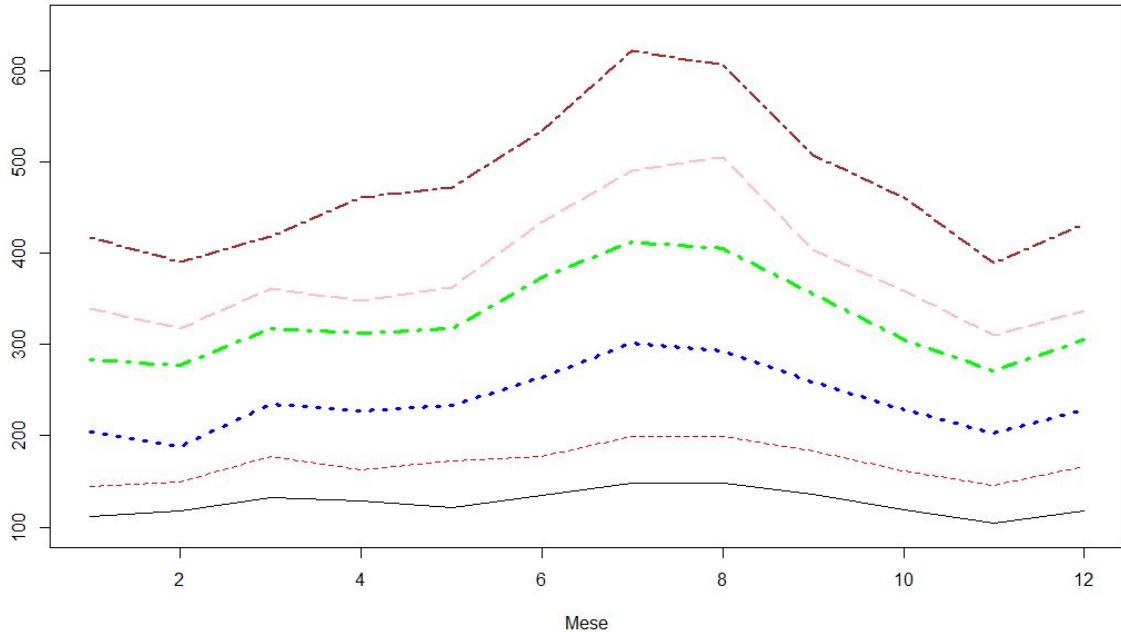


Il parametro **set.lags** viene impostato per settare i valori del ritardo  $k$ , in corrispondenza dei quali calcolare il grafico di autodispersione. Avendo impostato un intervallo per  $k$  che va da 1 a 12, vengono tracciati 12 grafici, ciascuno in corrispondenza di un valore di  $k$ . Un ultimo strumento per studiare la stagionalità di una serie storica è il seasonal plot, ossia il grafico della serie limitata ad un singolo anno, che permette di analizzare il comportamento in un anno (ovviamente tale grafico ha senso solo se i dati che si analizzano sono relativi a periodi inferiori all'anno). Questo grafico viene realizzato estrapolando manualmente, in quanto R non mette a disposizione un comando apposito per il seasonal plot, i dati relativi ad uno stesso anno e tracciandone il grafico. Di seguito tale operazione viene fatta solo per alcuni anni tra quelli presenti (i dati della serie vanno dal 1949 al 1960). Le serie estrapolate sono costituite da 12 osservazioni ciascuna e sono relative agli anni a partire dal 1949, con cadenza triennale: 1949, 1952, 1955, 1958.

```
> 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))
```

Per quanto riguarda il plotting, sono stati sovrapposti i grafici delle diverse serie annuali estrapolate, nel modo seguente:

```
> 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=3, col="blue", lwd=3)
> lines(x, y4, type="l", lty=4, col="green", lwd=3)
> lines(x, y5, type="l", lty=5, col="pink", lwd=2)
> lines(x, y6, type="l", lty=6, col="brown", lwd=2)
```



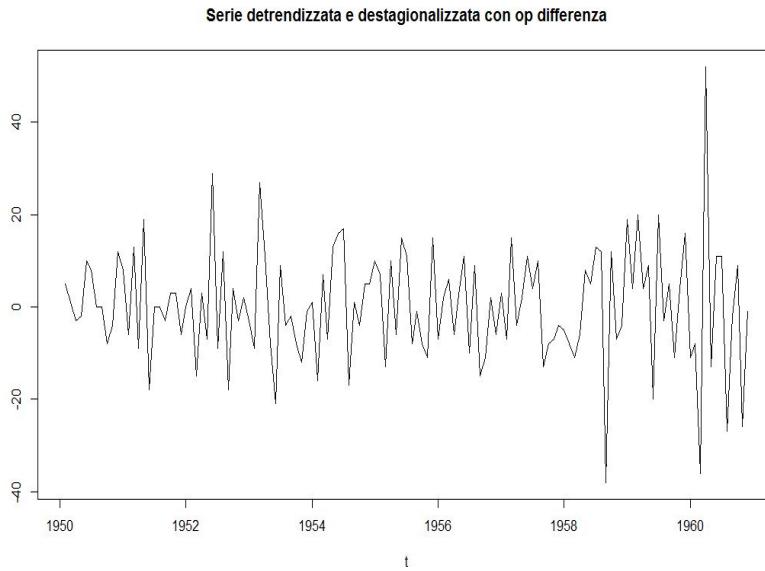
Dagli andamenti riportati in figura è possibile dedurre che la serie storica complessiva presenta una componente preponderante di stagionalità, in quanto, nei diversi anni analizzati, è presente un picco ogni anno in corrispondenza dei medesimi mesi. Inoltre, le diverse curve si assestano su livelli più alti (valori in ordinata più alti) e presentano una maggiore variabilità nei valori assunti dalla serie, il che implica la presenza di una componente di trend nella stagionalità.

Dalle considerazioni fatte è possibile concludere che la differenziazione più adatta è quella  $\nabla_1$  e  $\nabla_{12}$ , in quanto il primo operatore permette di detrendizzare la serie complessiva (rimuove la caratteristica di avere curve su livelli maggiori), mentre il secondo destagionalizza la serie (rimuove la caratteristica di avere picchi in presenza dello stesso periodo e andamento simile delle diverse curve). L'applicazione di questi due operatori si traduce nell'ottenimento della seguente serie:

$$Z_t = \nabla \nabla_{12} X_t = (1 - B)(1 - B^{12})X_t$$

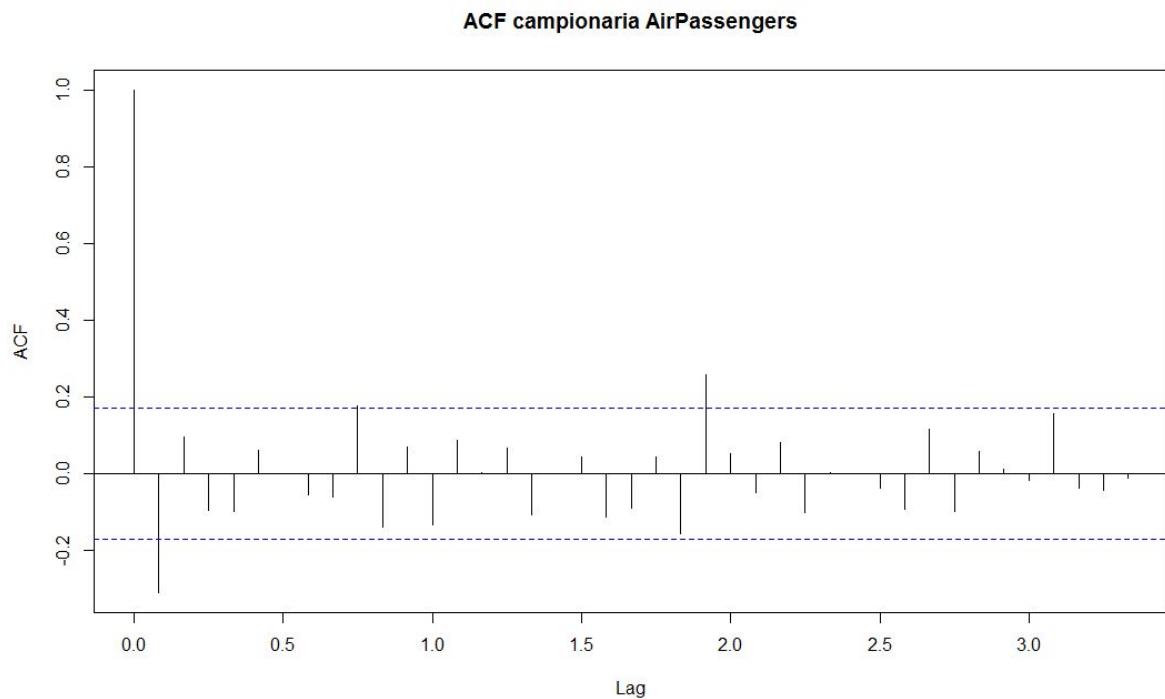
I comandi in R corrispondenti sono:

```
> yt <- diff(AirPassengers, lag=1, differences=1)
> zt <- diff(yt, lag=12, differences=1)
> plot(zt, ylab=" ", xlab="t", main="Serie detrendizzata e destagionalizzata con op differenza")
```



Una volta detrendizzata e destagionalizzata la serie, possiamo andare a esaminare l'eventuale correlazione esistente tra i dati, tramite l'andamento dell'autocorrelazione campionaria.

```
> acf(zt,lag.max = 40, main="ACF campionaria AirPassengers")
```

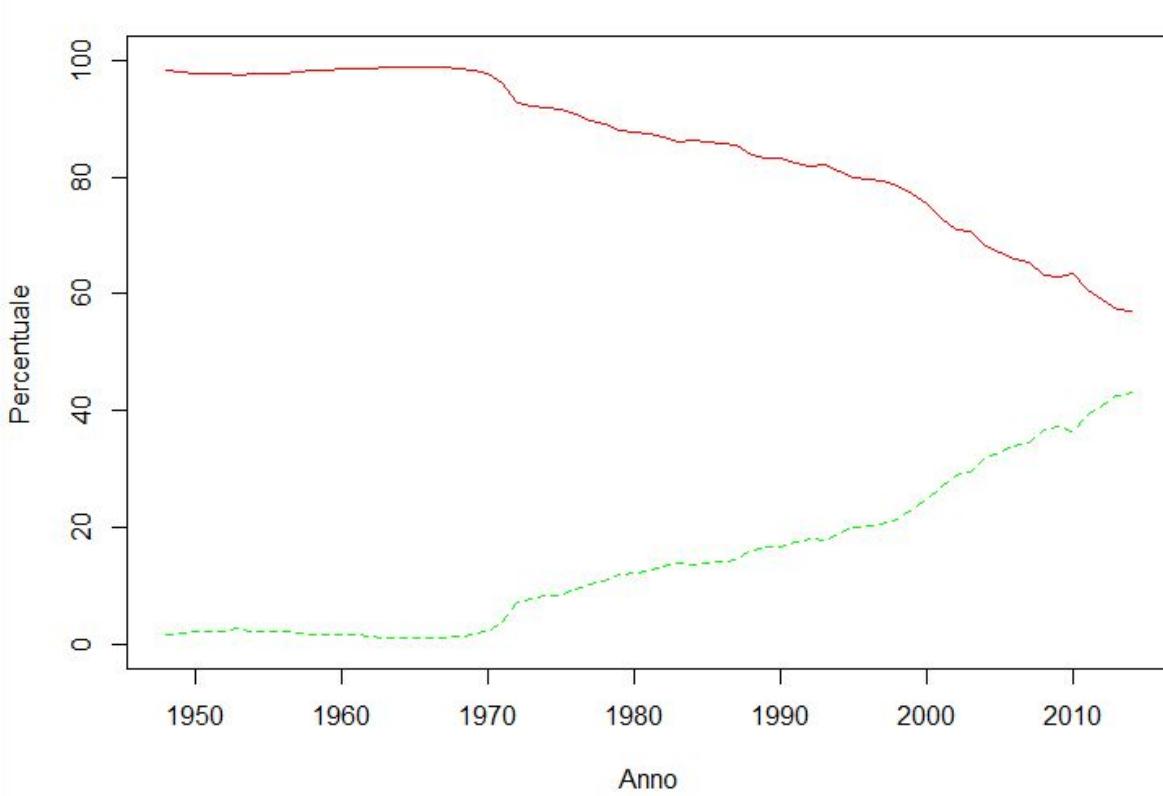


Come si evince dalla figura e come ci aspetteremmo, il processo non è riconducibile nettamente ad una delle categorie di processi stazionari studiati (si noti che alcuni valori dell'acf escono fuori dalla fascia blu evidenziata), ma potrebbe comunque avvicinarvisi. Serie Storiche.

## Matrimoni secondo il rito - Anni 1948-2014

Questa serie riporta le percentuali di matrimoni secondo il rito religioso o civile. Si è ritenuto opportuno selezionare i dati dal 1948 al 2014 escludendo i dati precedenti al 1948 in quanto dal 1943 al 1947 sono stati riportati solo i dati del totale dei matrimoni, quindi risulta impossibile stimare le percentuali dei singoli riti.

La somma delle percentuali di matrimoni effettuati secondo il rito civile e dei matrimoni effettuati secondo il rito religioso risulta corrispondere al 100% in tutti gli anni in analisi. Come primo passo è necessario tracciare il grafico dell'andamento della serie per le diverse categorie: dopo aver importato i dati tramite la funzione **read.csv2()**, sono state create le due serie temporali corrispondenti ai due tipi di dati (percentuale matrimoni secondo il rito civile, percentuale matrimoni secondo il rito religioso) e ne è stato fatto il plotting, come riportato di seguito.



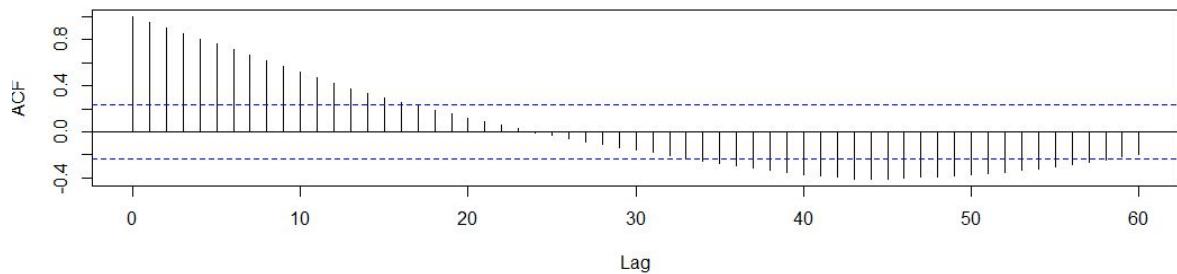
Dal grafico possiamo osservare che l'andamento dei matrimoni religiosi, rappresentato in rosso, tende a decrescere con l'avanzare degli anni. Mentre nei primi anni 50 era dominante la tradizione del matrimonio religiosa negli anni 2000 la tendenza del matrimonio secondo solo il rito civile, rappresentata in figura con il verde tratteggiato, è andato crescendo raggiungendo quasi le percentuali del matrimonio effettuato con il rito religioso.

Come prima operazione si vanno a tracciare gli andamenti dell'autocorrelazione campionaria globale e parziale per le serie viste. A monte di qualsiasi operazione di detrendizzazione ci si

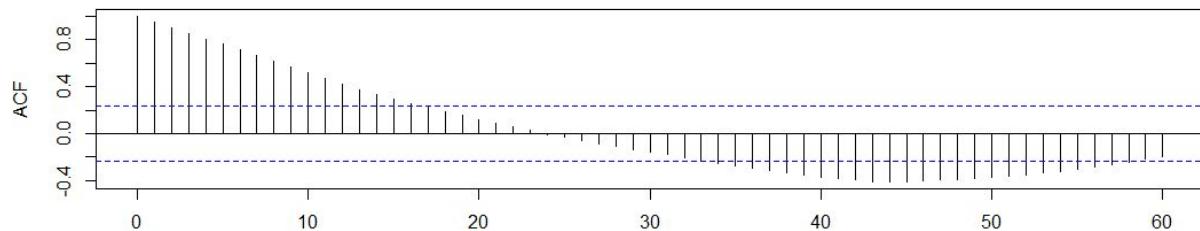
aspetta che la serie storica non presenti una ACF associata ad un qualche processo noto, men che meno ad un WN, in quanto la possibilità che essa derivi da un particolare processo stazionario è molto remota. Tramite i seguenti comandi tracciamo dunque gli andamenti desiderati:

```
> par(mfrow=c(2,1))
> acf(religiosi, lag.max =60, main="ACF campionaria percentuale religiosi")
> acf(civili, lag.max =60, main="ACF campionaria percentuale civili")
> par(mfrow=c(2,1))
> pacf(religiosi, lag.max =60, main="PACF campionaria percentuale religiosi")
> pacf(civili, lag.max =60, main="PACF campionaria percentuale civili")
```

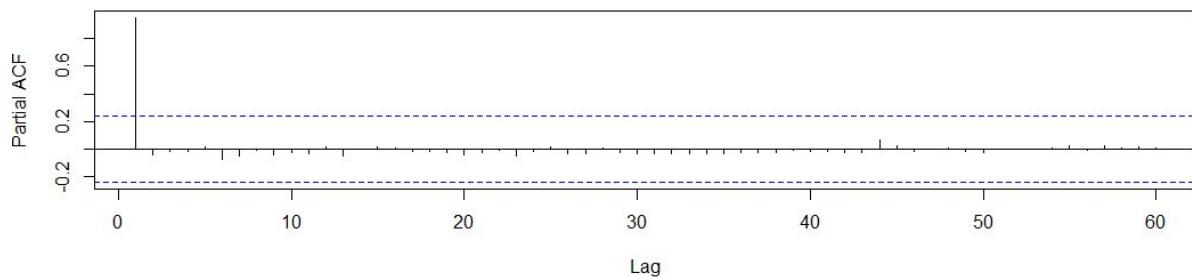
**ACF campionaria percentuale civili**

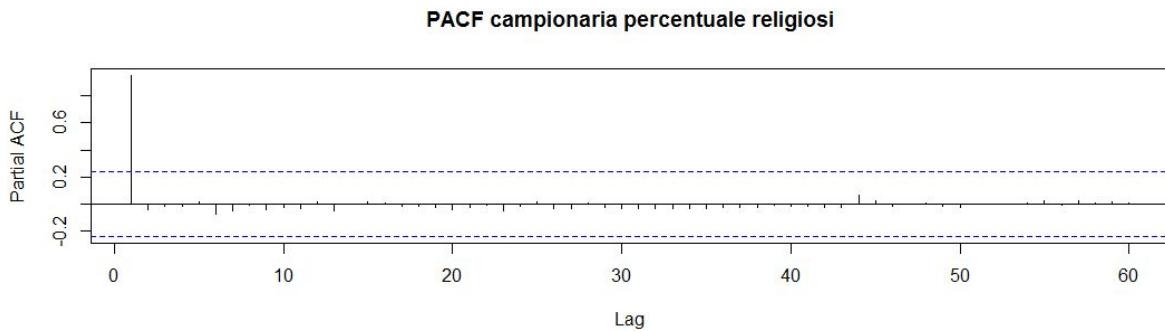


**ACF campionaria percentuale religiosi**



**PACF campionaria percentuale civili**





Dall'analisi è possibile dedurre che le serie considerate sono soggette a continue fluttuazioni attorno al trend principale che risulta essere quello di decrescita per il matrimonio secondo rito religioso e di crescita per matrimonio secondo rito civile. Questo provoca una decrescita delle funzioni di autocorrelazione globale fino al raggiungimento dello 0, per poi crescere nuovamente in modulo, ma su valori negativi.

Risulta allora evidente che i campioni di dati delle due serie non possono provenire da un processo di tipo WN, in quanto potrebbero non essere stazionari.

Tuttavia, può risultare interessante applicare un test in tal senso, per avere un'informazione quantitativa più precisa a riguardo. Inoltre, ciò consentirà di confrontare gli esiti del test prima e dopo le operazioni di detrendizzazione delle serie.

```
> Box.test(civts,type= "Ljung",lag=60)
  Box-Ljung test

data: civts
X-squared = 1302.6, df = 60, p-value < 2.2e-16

> Box.test(relts,type= "Ljung",lag=60)
  Box-Ljung test

data: relts
X-squared = 1302.6, df = 60, p-value < 2.2e-16
```

Il test di Ljung consiste in un test di ipotesi, basato su una statistica che, qualora l'ipotesi nulla fosse vera, risulta distribuita come una  $\chi^2$  con  $H$  gradi di libertà (parametro lag).

L'ipotesi nulla per tale test è che il campione di dati provenga da un processo WN e comunemente il livello di significatività del test è  $\alpha = 0,05$ .

Dal test appena effettuato è possibile osservare che il p-value restituito è un valore molto inferiore rispetto ad  $\alpha$ , quindi abbiamo un rigetto dell'ipotesi nulla come ci si aspettava dall'analisi dell'autocorrelazione e dall'andamento della serie.

Passiamo ora ad analizzare la serie detrendizzata: si stima il trend, supponendo un modello additivo, tramite media mobile semplice di ordine  $m = 1$ , in quanto i dati risultano essere

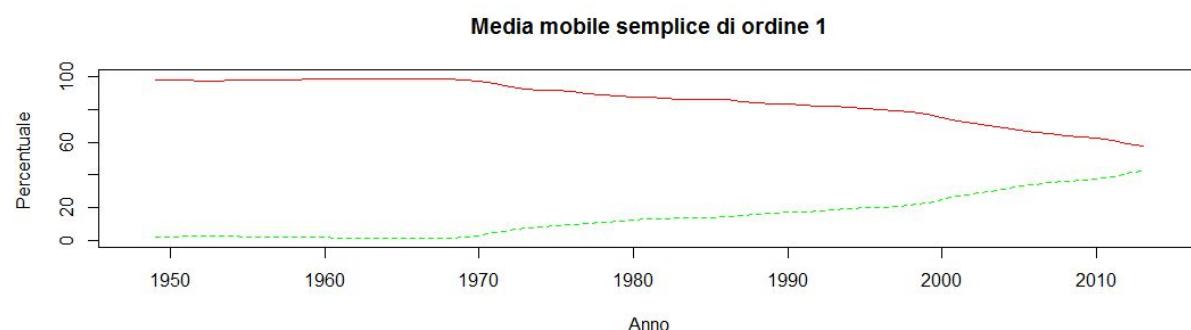
annuali. Un ordine basso per la media mobile risulta essere più fedele ai dati iniziali ma, di contro, riesce a smussare in misura minore eventuali irregolarità della serie. Questo non crea problemi in quanto i dati a disposizione hanno una granularità pari all'anno e non è possibile dunque analizzare la stagionalità (che avrebbe imposto un limite inferiore all'ordine della media mobile da usare). Essendo la media di ordine 1, il vettore dei pesi sarà  $\beta = (1/3, 1/3, 1/3)$ .

```
| > x=seq(1948,2014,1)
```

Ricordiamo che x rappresenta dunque la serie dal 1948 al 2014.

Per brevità rappresentiamo le due medie mobili ricavate, nel medesimo grafico.

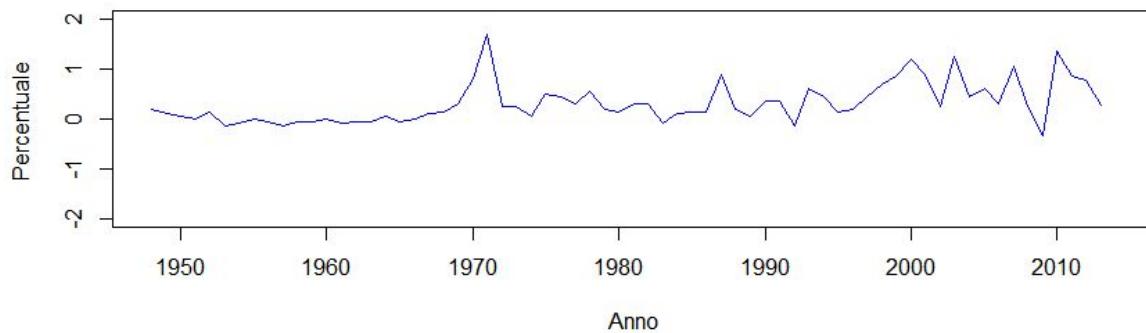
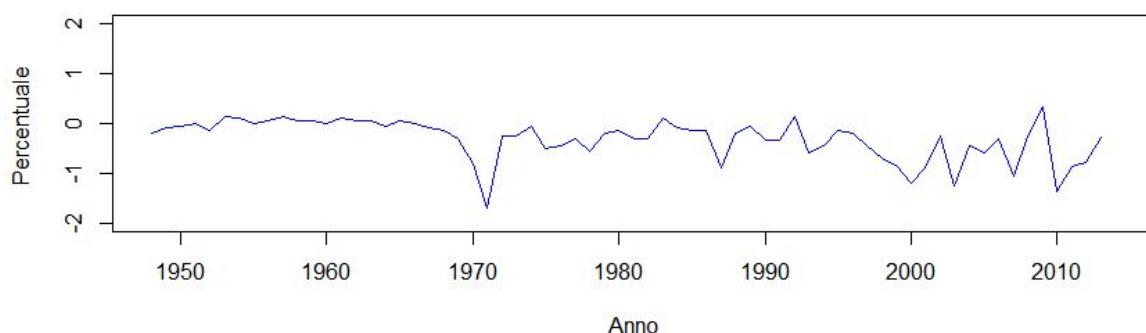
```
> beta<-rep(1/3,3)
> trend_religiosi<-filter(religiosi, beta, method = "convolution", sides=2, circular = FALSE)
> trend_civili<-filter(civili, beta, method = "convolution", sides=2, circular = FALSE)
> matplot(x, trend_religiosi,col = "red", type="l", ylim=c(0,100), ylab="Percentuale ", xlab="Anno", main="Media mobile semplice di ordine 1")
> lines(x, trend_civili, type="l", lty=2, col="green")
```



Rispetto all'andamento originale, i trend stimati risultano essere molto più smussati, quindi si suppone che siano dei buoni stimatori dell'andamento di fondo delle serie in esame.

Si procede, dunque, con la detrendizzazione vera e propria ottenendo i grafici desiderati attraverso i seguenti comandi:

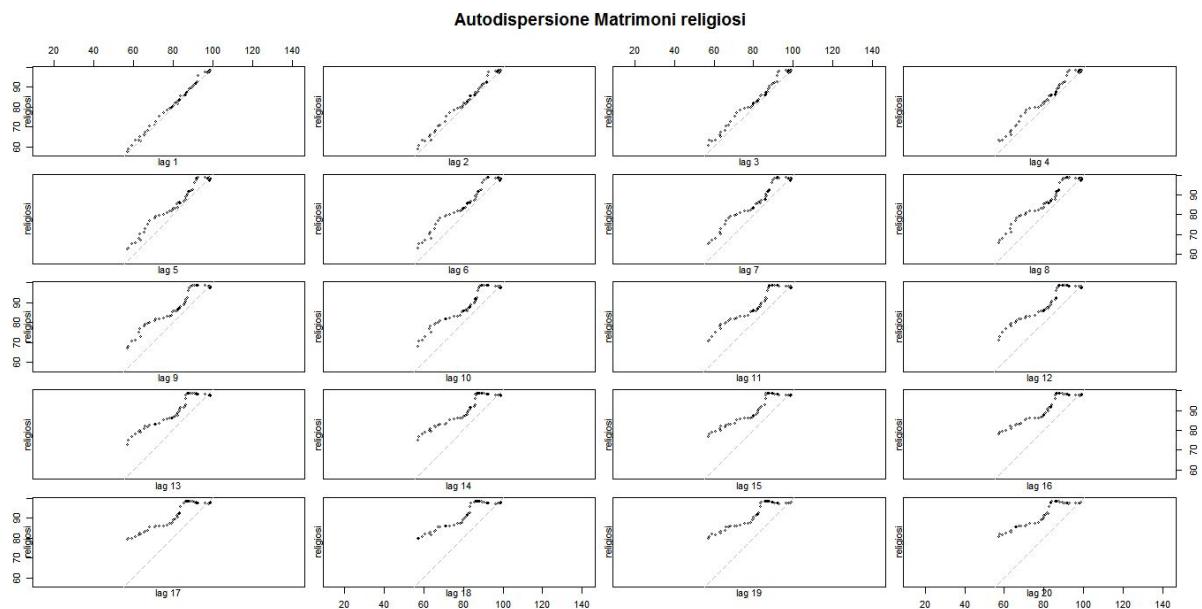
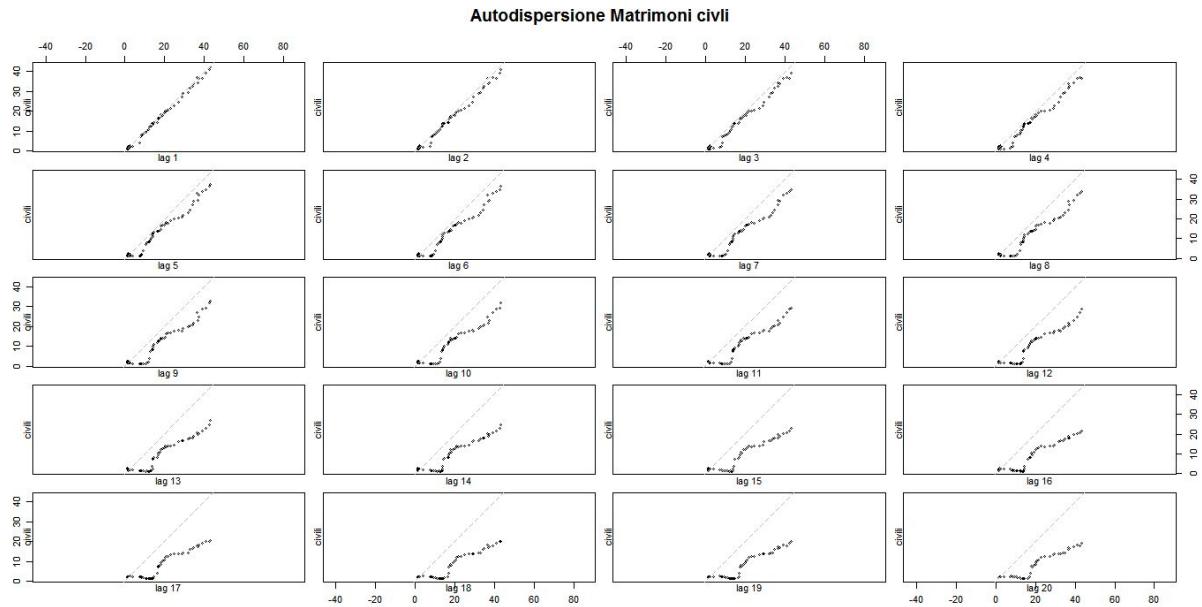
```
> detrend_religiosi<-religiosi-trend_religiosi
> detrend_civili<- civili-trend_civili
> par(mfrow=c(2,1))
> matplot(x, detrend_religiosi,col = "blue", type="l", ylim=c(-1,1), ylab="Percentuale ", xlab="Anno", main="Serie Matrimoni religiosi detrendizzata")
> matplot(x, detrend_civili,col = "blue", type="l", ylim=c(-1,1), ylab="Percentuale ", xlab="Anno", main="Serie Matrimoni civili detrendizzata")
```

**Serie Matrimoni religiosi detrendizzata****Serie Matrimoni civili detrendizzata**

Osservando i risultati le serie sono molto piatte il che implica che la componente preponderante delle serie è proprio il trend decrescente per il matrimonio secondo il rito religioso e quello crescente per il rito civile.

Infine, si procede ad analizzare i grafici di autodispersione delle serie considerate, al fine di confermare la presenza dei trend individuati.

```
> lag.plot(religiosi, set.lags = 1:20, type = "p", do.lines=FALSE, main = "Autodispersio
ne Matrimoni religiosi" )
> lag.plot(civili, set.lags = 1:20, type = "p", do.lines=FALSE, main = "Autodispersione
Matrimoni civili" )
.
```



I grafici di autodispersione evidenziano come, aumentando il lag, ossia il divario tra i valori rappresentati in ascissa e in ordinata, l'andamento dei dati si discosta da quello ideale (retta), per cui valori più distanti tenderanno sempre meno ad essere simili.

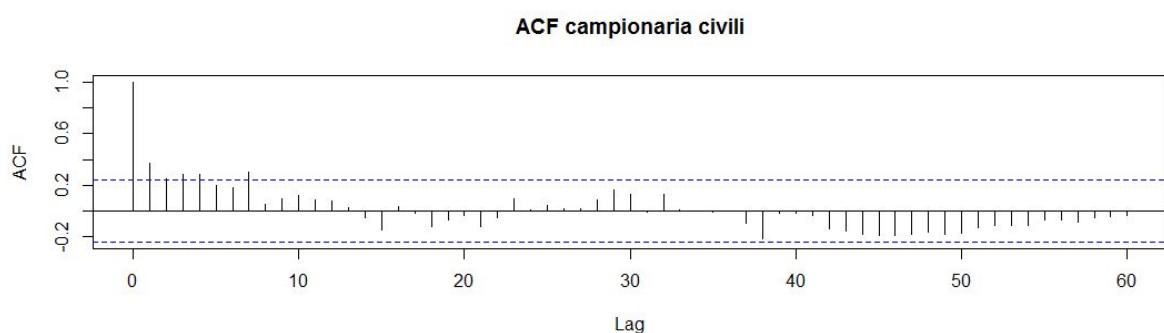
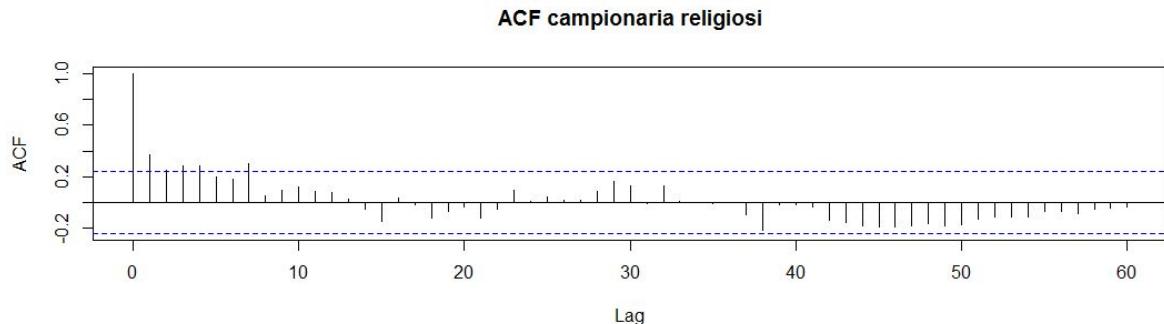
Infine, si effettua una detrendizzazione mediante operatore differenza e si ricalcolano le autocorrelazioni globali e parziali delle due serie.

```

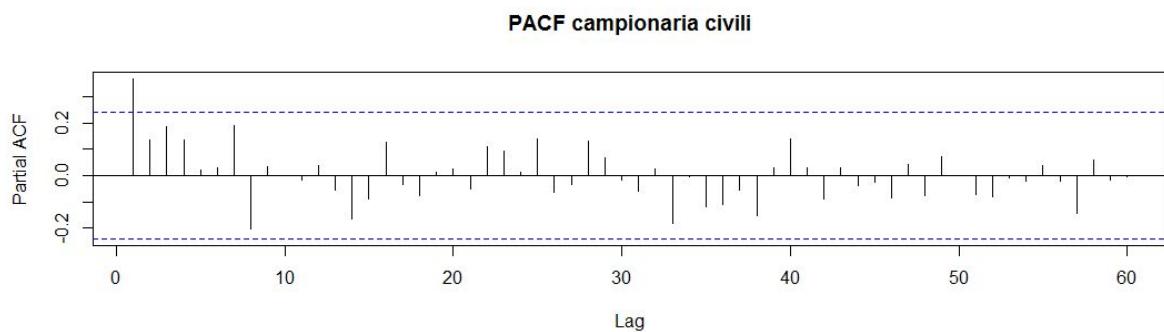
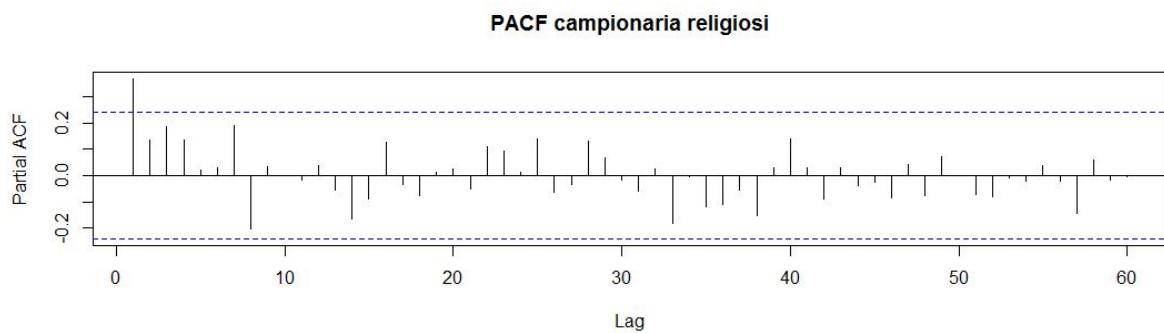
> diffreligiosi<-diff(religiosi, lag=1, differences = 1)
> diffcivili<-diff(civili, lag=1, differences = 1)
> par(mfrow=c(2,1))
> acf(diffreligiosi, lag.max = 60, main="ACF campionaria religiosi")
> acf(diffcivili, lag.max = 60, main="ACF campionaria civili")
> par(mfrow=c(2,1))
> pacf(diffreligiosi, lag.max = 60, main="PACF campionaria religiosi")
> pacf(diffcivili, lag.max = 60, main="PACF campionaria civili")

```

Ottenendo così prima le autocorrelazioni globali in ordine del matrimonio religioso e poi civile:



E infine nello stesso ordine le autocorrelazioni parziali:



Dai grafici possiamo osservare che con la rimozione del trend, principale causa di non stazionarietà, tramite differenziazione ci si avvicina maggiormente a quello che potrebbe essere un campione estratto da un processo stazionario.

E' interessante notare il risultato del test di Ljung sulle serie, dopo che queste sono detrendizzate, sia utilizzando la media mobile, sia usando l'operatore differenza.

```

> Box.test(detrend_religiosi, type = "Ljung", lag=60)
  Box-Ljung test

data: detrend_religiosi
X-squared = 80.057, df = 60, p-value = 0.04284

> Box.test(diffreligiosi, type = "Ljung", lag=60)
  Box-Ljung test

data: diffreligiosi
X-squared = 159.36, df = 60, p-value = 6.028e-11

> Box.test(detrend_civili, type = "Ljung", lag=60)
  Box-Ljung test

data: detrend_civili
X-squared = 80.057, df = 60, p-value = 0.04284

> Box.test(diffcivili, type = "Ljung", lag=60)
  Box-Ljung test

data: diffcivili
X-squared = 159.36, df = 60, p-value = 6.028e-11

```

Possiamo osservare dai test che effettuando una detrendizzazione e una differenziazione la serie non si avvicinerà ad essere una relazione di un processo WN poiché la probabilità che il residuo provenga da un WN non è sufficiente in particolare nel test di differenziazione risulta molto basso.

## Opere librerie pubblicate per materia trattata(1926-2014)

In questa sezione analizzeremo la serie riguardante la pubblicazioni delle opere librerie pubblicate per materia dal 1926 al 2014. Le materie trattate sono così divise:

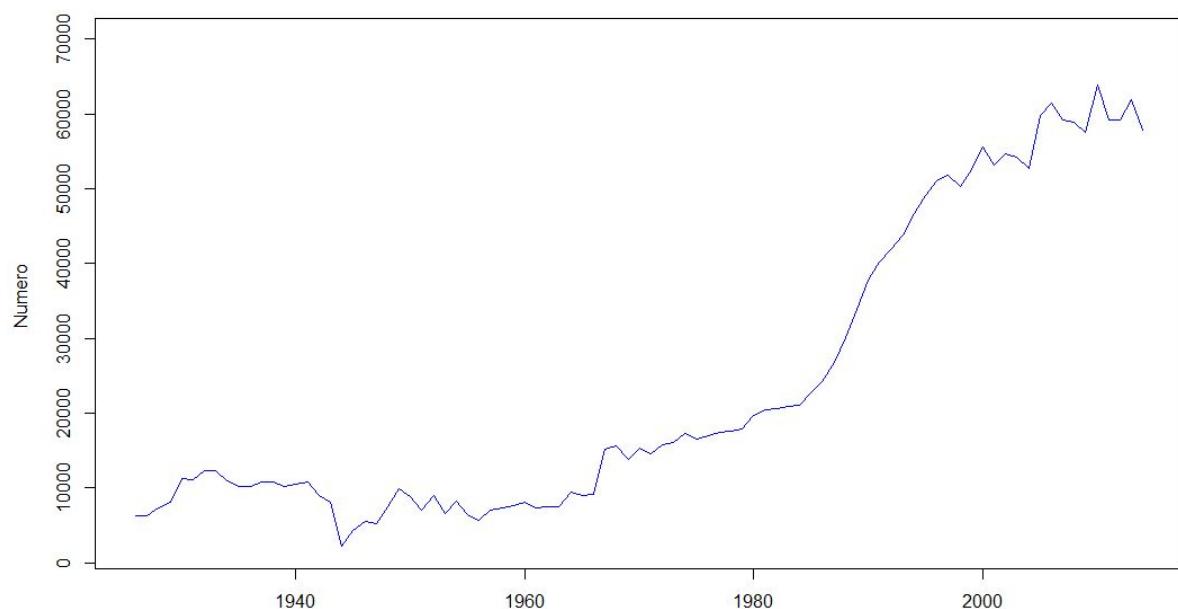
- Arti
- Letteratura
- Religione e teologia
- Scienze fisiche matematiche e naturali
- Scienze morali sociali e politiche
- Medicina
- Tecnologia
- Altro

Per prima cosa, dopo aver importato il dataset e associato le colonne agli opportuni campi, come fatto precedentemente, guardiamo prima il plot del totale delle pubblicazioni avvenute dal 1926 al 2014 e successivamente quelle delle pubblicazioni per materia trattata (in questo caso si è scelto di utilizzare per una prima impressione solo un plot).

```
> arti<-DatasetLibri$Arti
> letteratura<-DatasetLibri$Letteratura
> religione_e_teologia<-DatasetLibri$`Religione e teologia`
> scientifiche<-DatasetLibri$`Scienze fisiche matematiche e naturali`
> scientifiche2<-DatasetLibri$`Scienze morali sociali e politiche`
> medici<-DatasetLibri$`Medicina e igiene`
> tecnologia<-DatasetLibri$Tecnologia
> altro<-DatasetLibri$Altro
> tot<-DatasetLibri$Tot

> x=seq(1926,2014,1)

> matplot(x, tot, type="l", ylim=c(2000,70000), ylab="Numero ", xlab="Anno", col="blue")
```



Possiamo vedere dal totale delle opere librarie pubblicate un andamento crescente anche se non continuo nel tempo. Per prima cosa è importante notare che la fonte dei dati ci informa che dal 1951 è possibile il confronto tra i dati riferiti ai diversi anni mentre non sono direttamente comparabili con quelli rilevati nel periodo prebellico.

Inoltre dal 1964 sono state incluse tra le opere librarie in analisi anche quelle di 50 e 100 pagine e dal 1967 anche le ristampe e gli opuscoli da 5 a 48 pagine influenzando così la crescita.

Dalle osservazioni precedente, visto che i dati precedenti al 1951 non sono comparabili selezioneremo un sottoinsieme dei dati partendo in esame dall'anno postbellico 1951.

```
> x=seq(1951,2014,1)
```

```

> data2<-DatasetLibri[which(DatasetLibri$Anni > 1950),]
> arti<-data2$Arti
> letteratura<-data2$Letteratura
> religione_e_teologia<-data2$`Religione e teologia`
> scientifiche<-data2$`Scienze fisiche matematiche e naturali`
> scientifiche2<-data2$`Scienze morali sociali e politiche`
> medici<-data2$`Medicina e igiene`
> tecnologia<-data2$Tecnologia
> altro<-data2$Altro

```

In *data2* ora sono presenti tutti i dati che andremo ad analizzare.

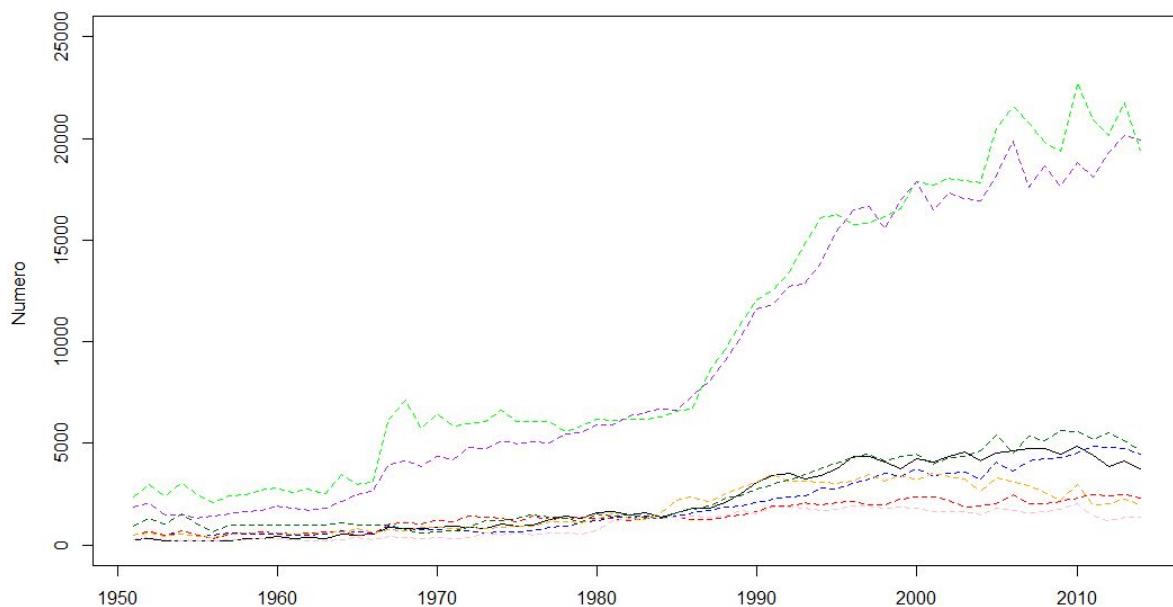
Mostriamo dunque inizialmente in unico plot tutte le serie e successivamente le singole dove sull'asse delle ascisse riportiamo gli anni e sull'asse delle ordinate il numero di pubblicazioni.

```

> matplot(x, arti, type="l", ylim=c(0,30000), ylab="Numero ", xlab="Anno")
> lines(x, letteratura, type="l", lty=2, col="green")
> lines(x, religione_e_teologia, type="l", lty=2, col="blue")
> lines(x, scientifiche, type="l", lty=2, col="red")
> lines(x, scientifiche2, type="l", lty=2, col="purple")
> lines(x, medici, type="l", lty=2, col="pink")
> lines(x, tecnologia, type="l", lty=2, col="orange")
> lines(x, altro, type="l", lty=2, col="darkgreen")

```

Dal codice osserviamo che le linee in verde rappresentano le materie di letteratura, in blu quelle di religione e teologia, in rosso quelle scientifiche matematiche e naturali, in viola le scientifiche morali politiche e sociali, in rosa quelle mediche, in arancione quelle tecnologiche, in verde scuro tutte le altre e infine di default quelle artistiche mostrate con il colore nero. Inoltre per comodità le serie riguardanti le materie Scienze fisiche matematiche e naturali e Scienze morali sociali e politiche sono state nominate scientifiche e scientifiche2.



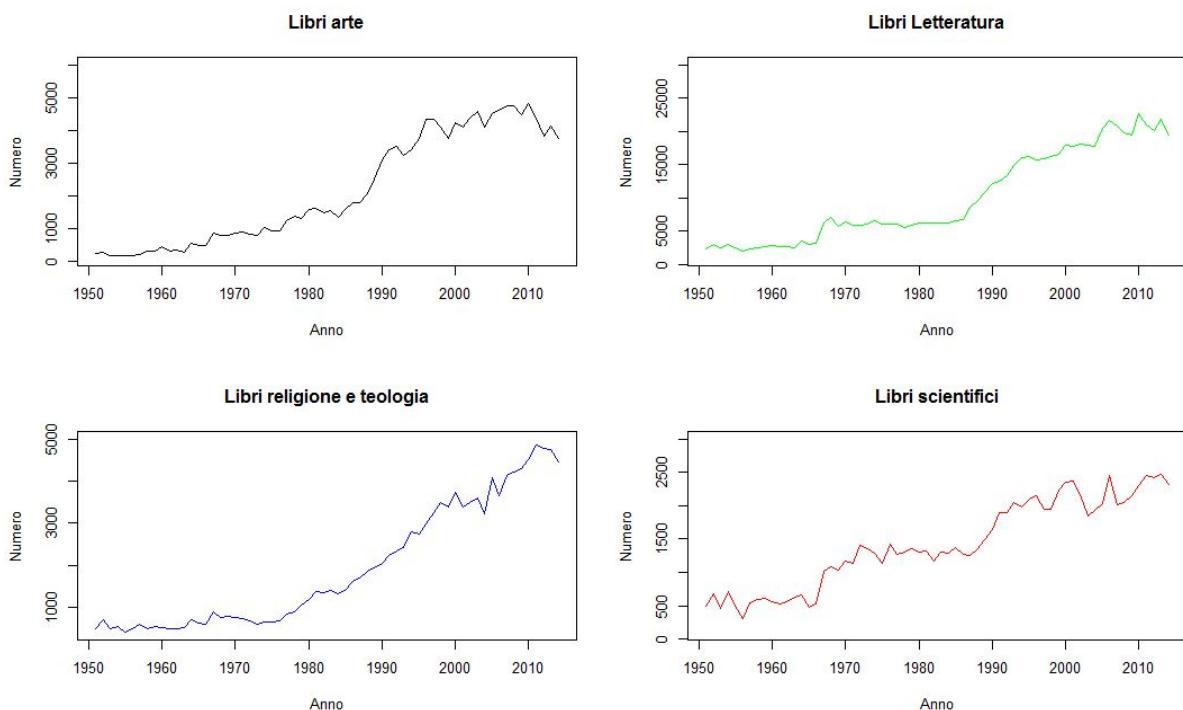
Non è visibilmente chiaro il risultato ma da una prima analisi possiamo osservare che le opere librarie di natura letteraria e scientifica di tipo morali, politiche e sociali hanno un andamento assai più crescente rispetto le altre tipologie di opere librarie.

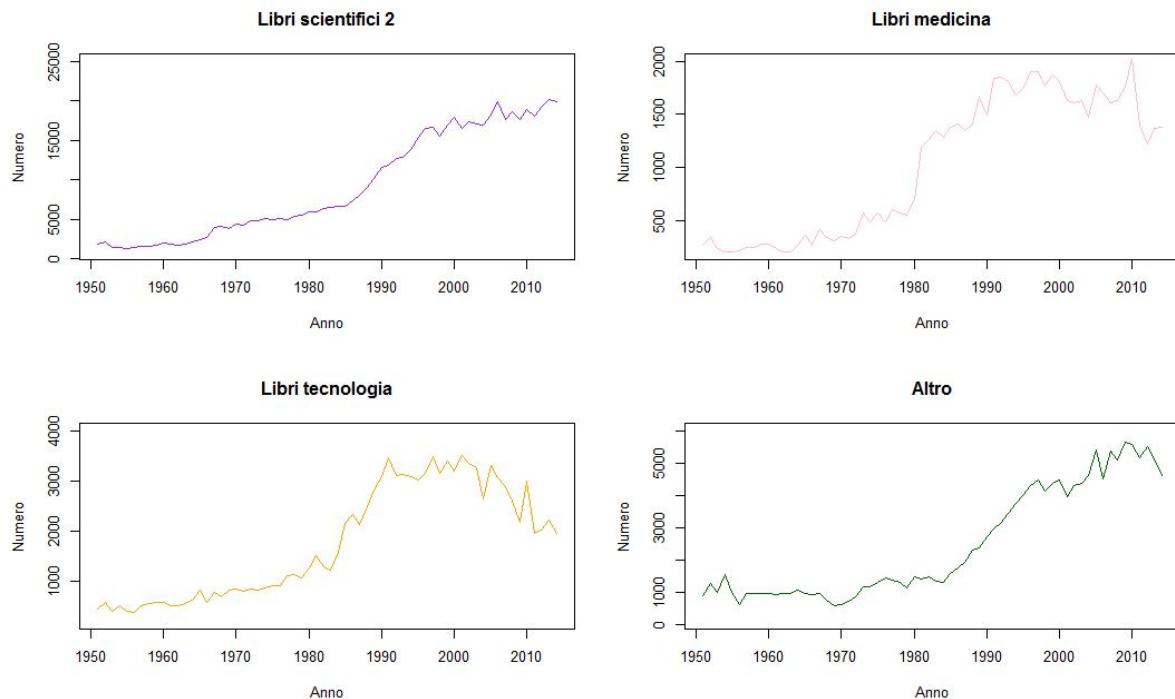
Mantenendo gli stessi colori abbiamo:

```

> par(mfrow=c(2,2))
> range(arti)
[1] 168 4838
> matplot(x, arti, type = "l", ylim=c(100,6000), ylab = "Numero", xlab = "Anno",
  main="Libri arte", col="black")
> range(letteratura)
[1] 2056 22740
> matplot(x, letteratura, type = "l", ylim=c(900,30000), ylab = "Numero", xlab =
  "Anno", main="Libri Letteratura", col="green")
> range(religione_e_teologia)
[1] 417 4849
> matplot(x, religione_e_teologia, type = "l", ylim=c(400,5000), ylab = "Numero"
, xlab = "Anno", main="Libri religione e teologia", col="blue")
> range(scientifiche)
[1] 306 2473
> matplot(x, scientifiche, type = "l", ylim=c(100,3000), ylab = "Numero", xlab =
  "Anno", main="Libri scientifici", col="red")
> range(scientifiche2)
[1] 1284 20126
> matplot(x, scientifiche2, type = "l", ylim=c(800,25000), ylab = "Numero", xlab =
  "Anno", main="Libri scientifici 2", col="purple")
> range(medici)
[1] 203 2025
> matplot(x, medici, type = "l", ylim=c(200,2000), ylab = "Numero", xlab = "Anno",
  main="Libri medicina", col="pink")
> range(tecnologia)
[1] 387 3506
> matplot(x, tecnologia, type = "l", ylim=c(200,4000), ylab = "Numero", xlab = "A
nno", main="Libri tecnologia", col="orange")
> range(altra)
[1] 605 5636
> matplot(x, altra, type = "l", ylim=c(100,6000), ylab = "Numero", xlab = "Anno"
, main="Altro", col="darkgreen")

```





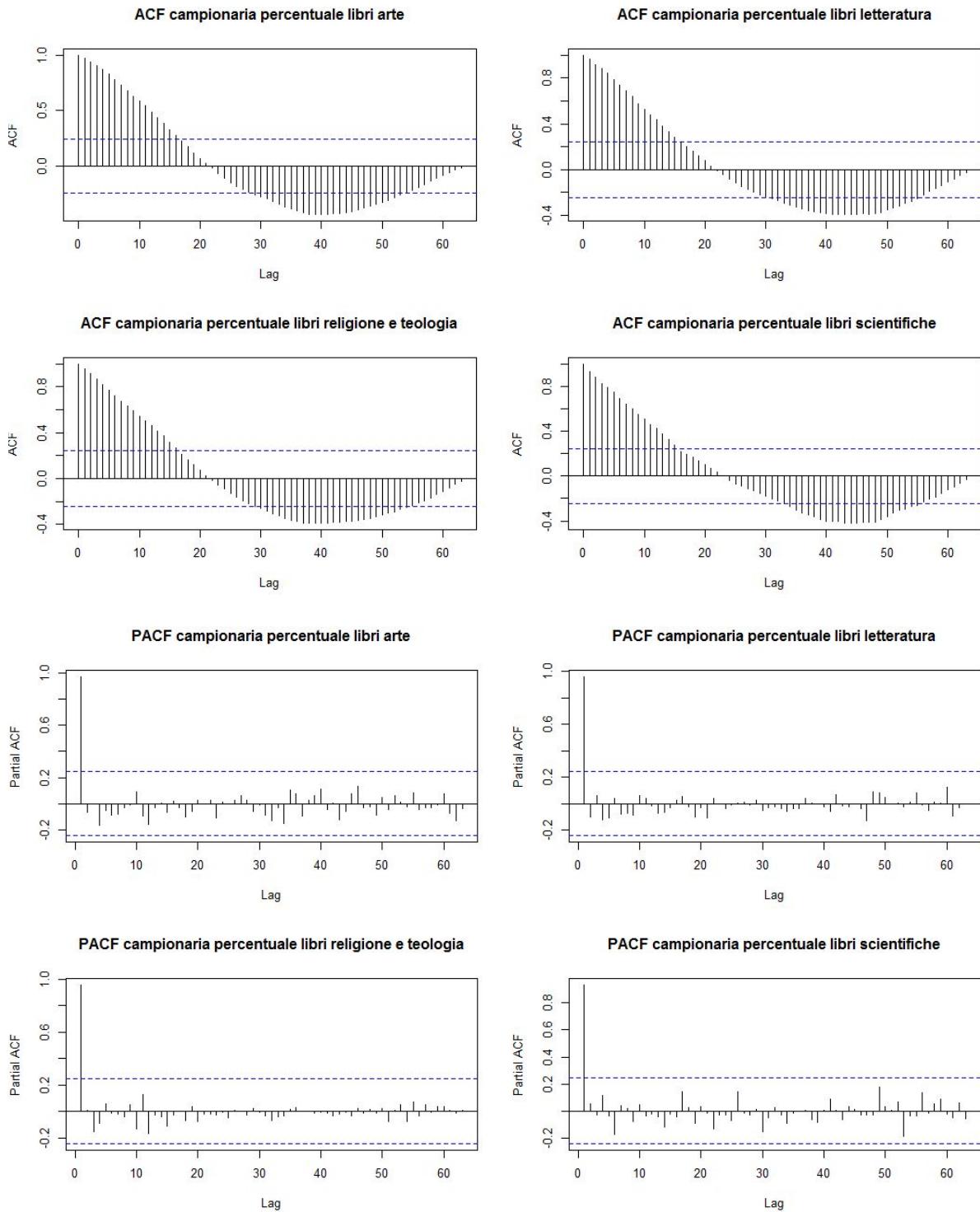
Si può notare un comportamento più o meno crescente fino al 2000 per tutti i numeri di pubblicazione delle opere librarie, successivamente le opere librarie in particolare di materie tecnologica hanno subito varie fluttuazione.

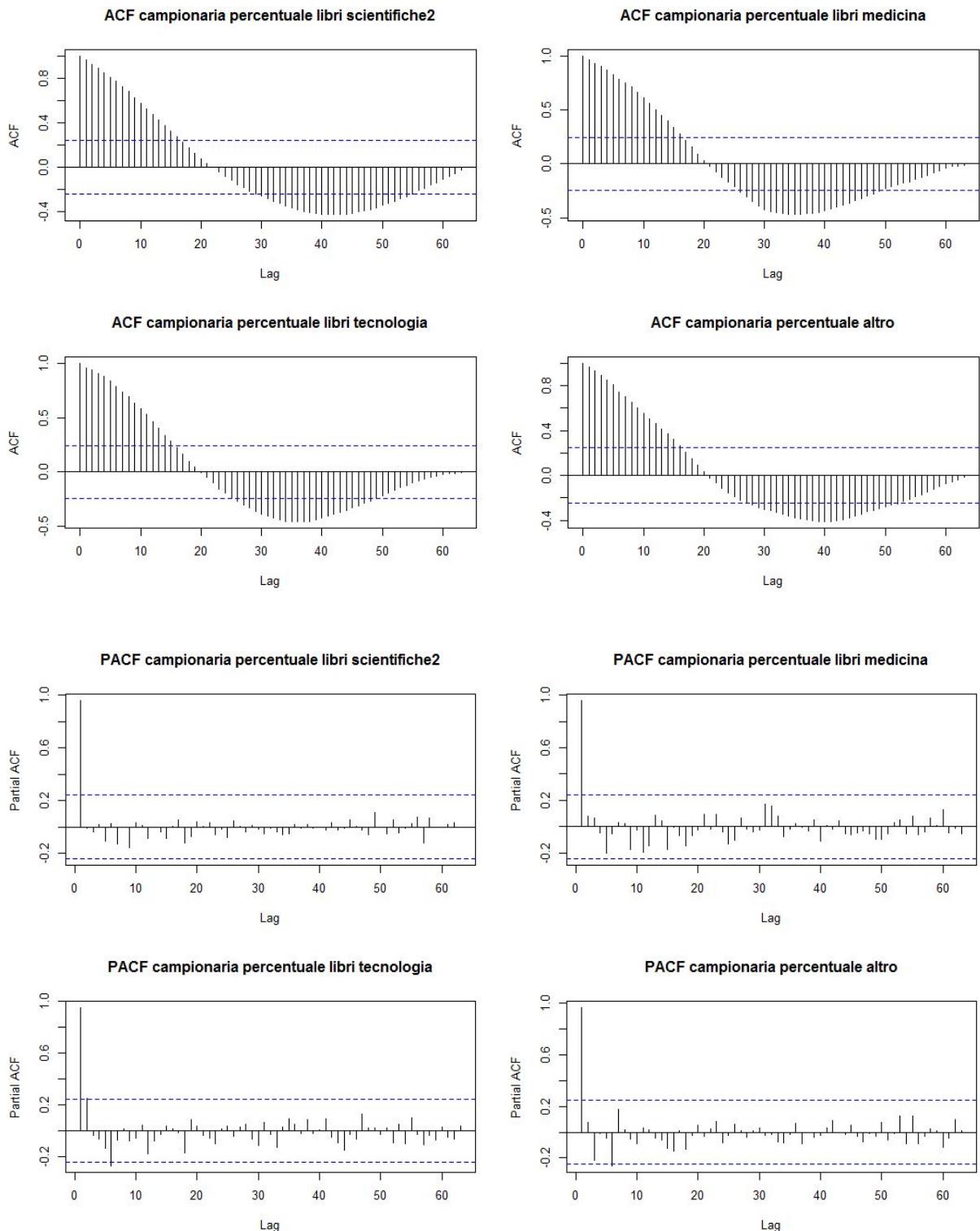
Si procede ora con la rappresentazione della acf campionaria e dell'autocorrelazione parziale.

```

> par(mfrow=c(2,2))
> acf(arti, lag.max =100, main="ACF campionaria percentuale libri arte")
> acf(letteratura, lag.max =100, main="ACF campionaria percentuale libri letteratura")
> acf(religione_e_teologia, lag.max =100, main="ACF campionaria percentuale libri religione e teologia")
> acf(scientifiche, lag.max =100, main="ACF campionaria percentuale libri scientifiche")
> acf(scientifiche2, lag.max =100, main="ACF campionaria percentuale libri scientifiche2")
> acf(medici, lag.max =100, main="ACF campionaria percentuale libri medicina")
> acf(tecnologia, lag.max =100, main="ACF campionaria percentuale libri tecnologia")
> acf(altri, lag.max =100, main="ACF campionaria percentuale altro")
> par(mfrow=c(2,2))
> pacf(arti, lag.max =100, main="PACF campionaria percentuale libri arte")
> pacf(letteratura, lag.max =100, main="PACF campionaria percentuale libri letteratura")
> pacf(religione_e_teologia, lag.max =100, main="PACF campionaria percentuale libri religione e teologia")
> pacf(scientifiche, lag.max =100, main="PACF campionaria percentuale libri scientifiche")
> pacf(scientifiche2, lag.max =100, main="PACF campionaria percentuale libri scientifiche2")
> pacf(medici, lag.max =100, main="PACF campionaria percentuale libri medicina")
> pacf(tecnologia, lag.max =100, main="PACF campionaria percentuale libri tecnologia")
> pacf(altri, lag.max =100, main="PACF campionaria percentuale altro")

```





Di seguito come fatto nella precedente sezione, si riportano i risultati del test di Ljung effettuato sulle serie prima della detrendizzazione. I valori sono tutti a favore del fatto che nessuno dei campioni esaminati, probabilmente non stazionari, può provenire da un processo WN.

```

> Box.test(arti,type="Ljung", lag=50)
  Box-Ljung test

data: arti
X-squared = 1183.1, df = 50, p-value < 2.2e-16

> Box.test(letteratura,type="Ljung", lag=50)
  Box-Ljung test

data: letteratura
X-squared = 1036.4, df = 50, p-value < 2.2e-16

> Box.test(religione_e_teologia,type="Ljung", lag=50)
  Box-Ljung test

data: religione_e_teologia
X-squared = 1028.3, df = 50, p-value < 2.2e-16

> Box.test(scientifiche,type="Ljung", lag=50)
  Box-Ljung test

data: scientifiche
X-squared = 997.69, df = 50, p-value < 2.2e-16

> Box.test(scientifiche2,type="Ljung", lag=50)
  Box-Ljung test

data: scientifiche2
X-squared = 1145.6, df = 50, p-value < 2.2e-16

> Box.test(medici,type="Ljung", lag=50)
  Box-Ljung test

data: medici
X-squared = 1207.4, df = 50, p-value < 2.2e-16

> Box.test(tecnoologia,type="Ljung", lag=50)
  Box-Ljung test

data: tecnoologia
X-squared = 1139.8, df = 50, p-value < 2.2e-16

> Box.test(altro,type="Ljung", lag=50)
  Box-Ljung test

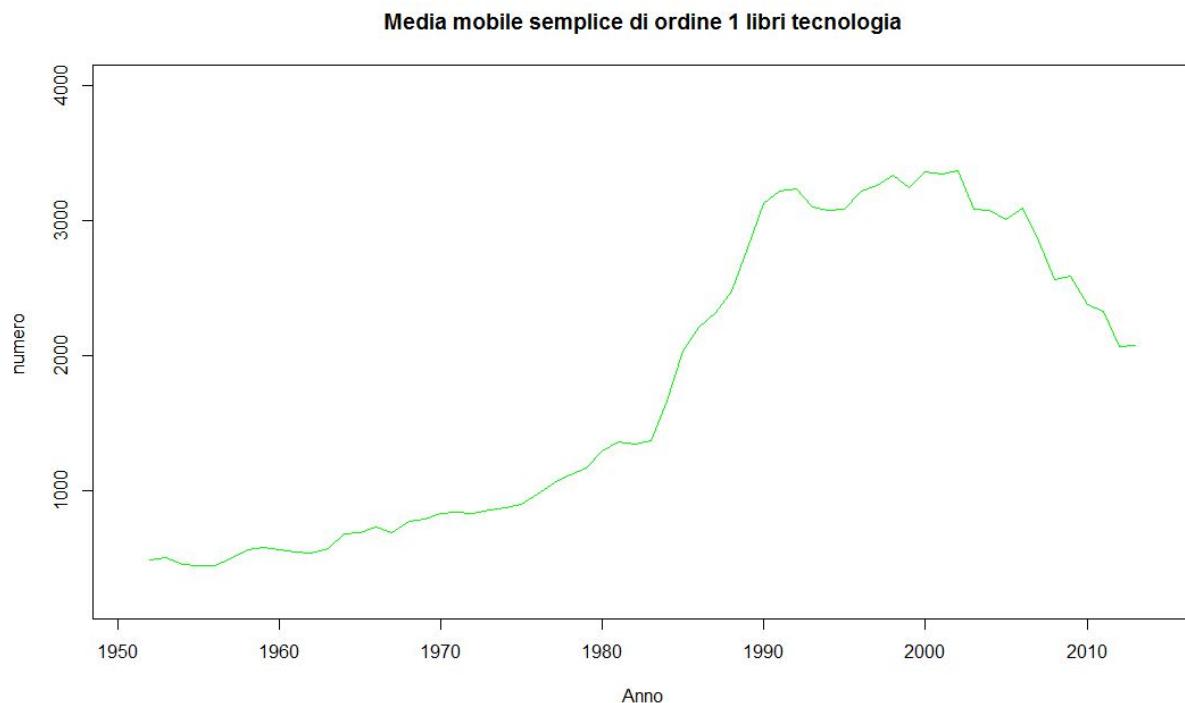
data: altro
X-squared = 1064.7, df = 50, p-value < 2.2e-16

```

Di seguito si riporta la stima del trend delle diverse serie, fatta tramite una media mobile semplice di ordine 2. Tali dati sono annuali e non è possibile dunque individuare eventuali stagionalità. Si riesce quindi a stimarne il trend, mantenendo tuttavia una certa fedeltà ai dati originali. Di seguito si riporta anche un confronto tra una media mobile di ordine 1 e una

media mobile di ordine 2, fatte sulla serie riguardante la pubblicazione di opere librerie di natura tecnologica ( basta applicare alla serie la funzione filter(), con un vettore dei pesi  $\alpha=(1/5,1/5,1/5,1/5,1/5)$ ).

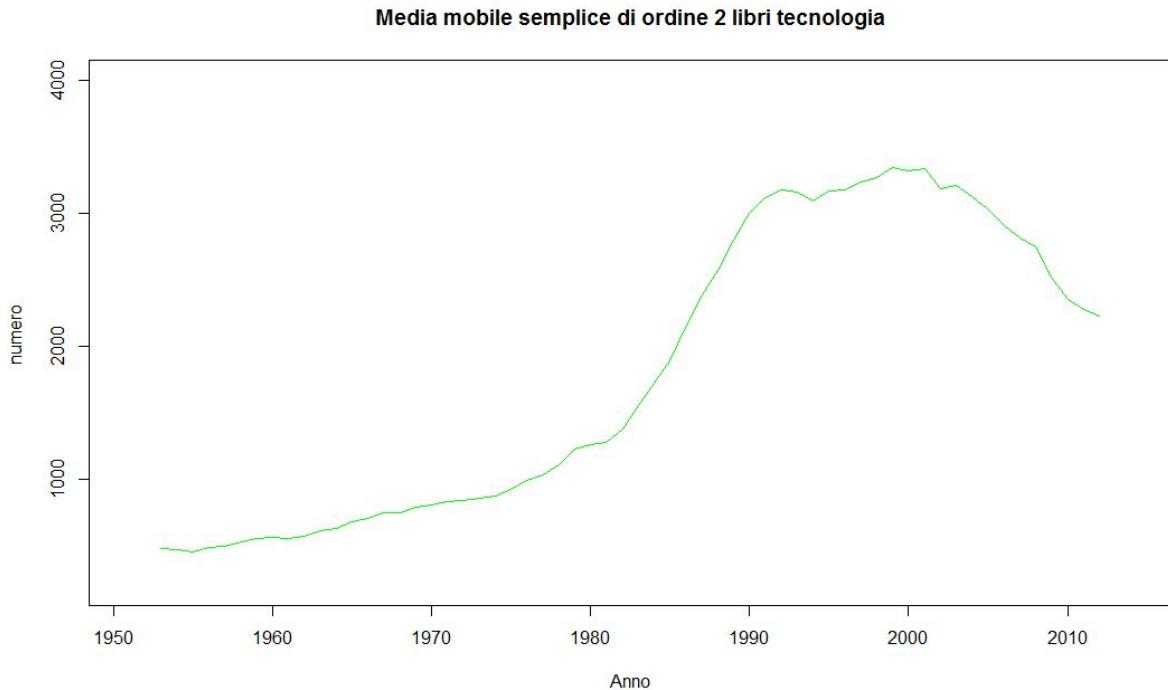
```
> beta<-rep(1/3,3)
> trend_tech<-filter(tecno, beta, method = "convolution", sides=2, circular = FALSE)
> matplot(x, trend_tech, col = "green", type="l", ylim=c(200,4000), ylab="numero ", xlab="Anno", main="Media mobile semplice di ordine 1 libri tecnologia")
```



```

> alpha<-rep(1/5,5)
> trend_tecnologia<-filter(tecnologia, alpha, method = "convolution", sides=2, circular = FALSE)
>
> matplot(x, trend_tecnologia,col = "green", type="l", ylim=c(200,4000), ylab="numero ", xlab="Anno",
  main="Media mobile semplice di ordine 2 libri tecnologia")

```



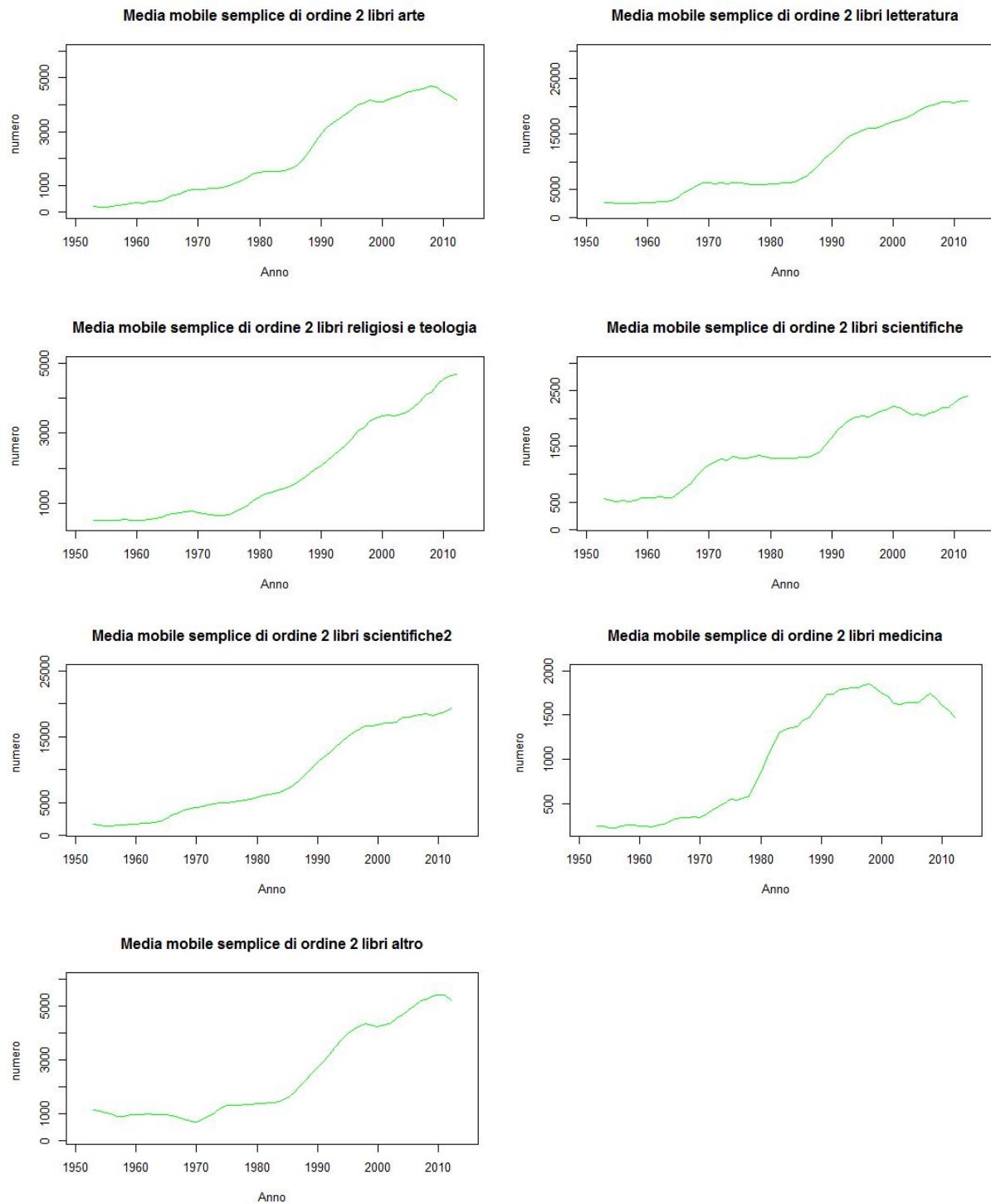
Da tale grafico possiamo osservare chiaramente come il trend sia meno smussato nelle sue irregolarità effettuando una media mobile semplice di ordine 2.

In questo caso, si andrà ad effettuare la detrendizzazione delle serie considerate utilizzando una media mobile di ordine 2. Si prosegue con le altre serie.

```

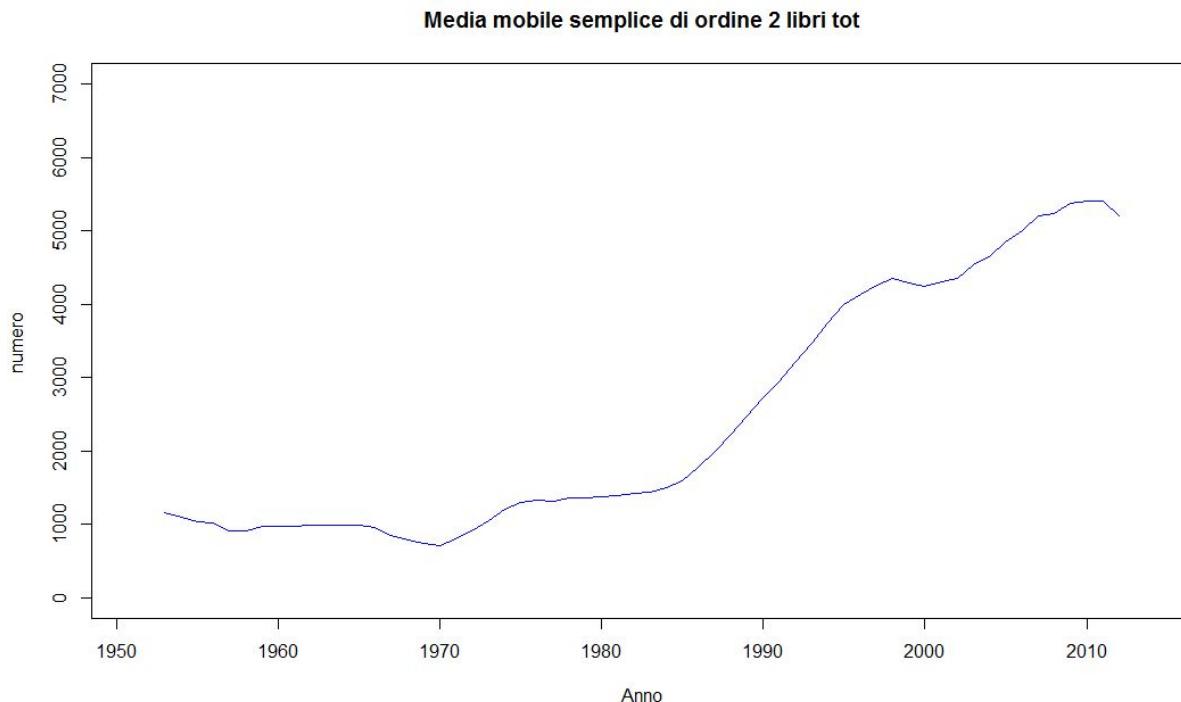
> trend_arte<-filter(arti, alpha, method = "convolution", sides=2, circular = FALSE)
>
> matplot(x, trend_arte,col = "green", type="l", ylim=c(0,6000), ylab="numero ", xlab="Anno", main="Media mobile semplice di ordine 1 libri arte")
>
> trend_letteratura<-filter(letteratura, alpha, method = "convolution", sides=2, circular = FALSE)
> trend_religioseteologia<-filter(religione_e_teologia, alpha, method = "convolution", sides=2, circular = FALSE)
> trend_scientifiche<-filter(scientifiche, alpha, method = "convolution", sides=2, circular = FALSE)
> trend_scientifiche2<-filter(scientifiche2, alpha, method = "convolution", sides=2, circular = FALSE)
> trend_medicina<-filter(medici, alpha, method = "convolution", sides=2, circular = FALSE)
> trend_tecnologia<-filter(tecnologia, alpha, method = "convolution", sides=2, circular = FALSE)
> trend_alto<-filter(alto, alpha, method = "convolution", sides=2, circular = FALSE)
> matplot(x, trend_letteratura,col = "green", type="l", ylim=c(900,30000), ylab="numero ", xlab="Anno ", main="Media mobile semplice di ordine 2 libri letteratura")
> matplot(x, trend_religioseteologia,col = "green", type="l", ylim=c(400,5000), ylab="numero ", xlab="Anno ", main="Media mobile semplice di ordine 2 libri religiosi e teologia")
> matplot(x, trend_scientifiche,col = "green", type="l", ylim=c(100,3000), ylab="numero ", xlab="Anno ", main="Media mobile semplice di ordine 2 libri scientifiche")
> matplot(x, trend_scientifiche2,col = "green", type="l", ylim=c(800,25000), ylab="numero ", xlab="Anno ", main="Media mobile semplice di ordine 2 libri scientifiche2")
> matplot(x, trend_medicina,col = "green", type="l", ylim=c(200,2000), ylab="numero ", xlab="Anno ", main="Media mobile semplice di ordine 2 libri medicina")
> matplot(x, trend_tecnologia,col = "green", type="l", ylim=c(200,4000), ylab="numero ", xlab="Anno ", main="Media mobile semplice di ordine 2 libri tecnologia")
> matplot(x, trend_alto,col = "green", type="l", ylim=c(100,6000), ylab="numero ", xlab="Anno ", main="Media mobile semplice di ordine 2 libri altro")

```



Anche sul totale (a partire dal 1951 per i motivi sopra spiegati) viene analizzato effettuando una media mobile di ordine 2.

```
> tot2<-data2$Tot
> trend_tot<-filter(tot2, alpha, method = "convolution", sides=2, circular = FALSE)
> matplot(x, trend_ultimo,col = "blue", type="l", ylim=c(0,7000), ylab="numero ", xlab="Anno", main="Media mobile semplice di ordine 2 libri tot")
```

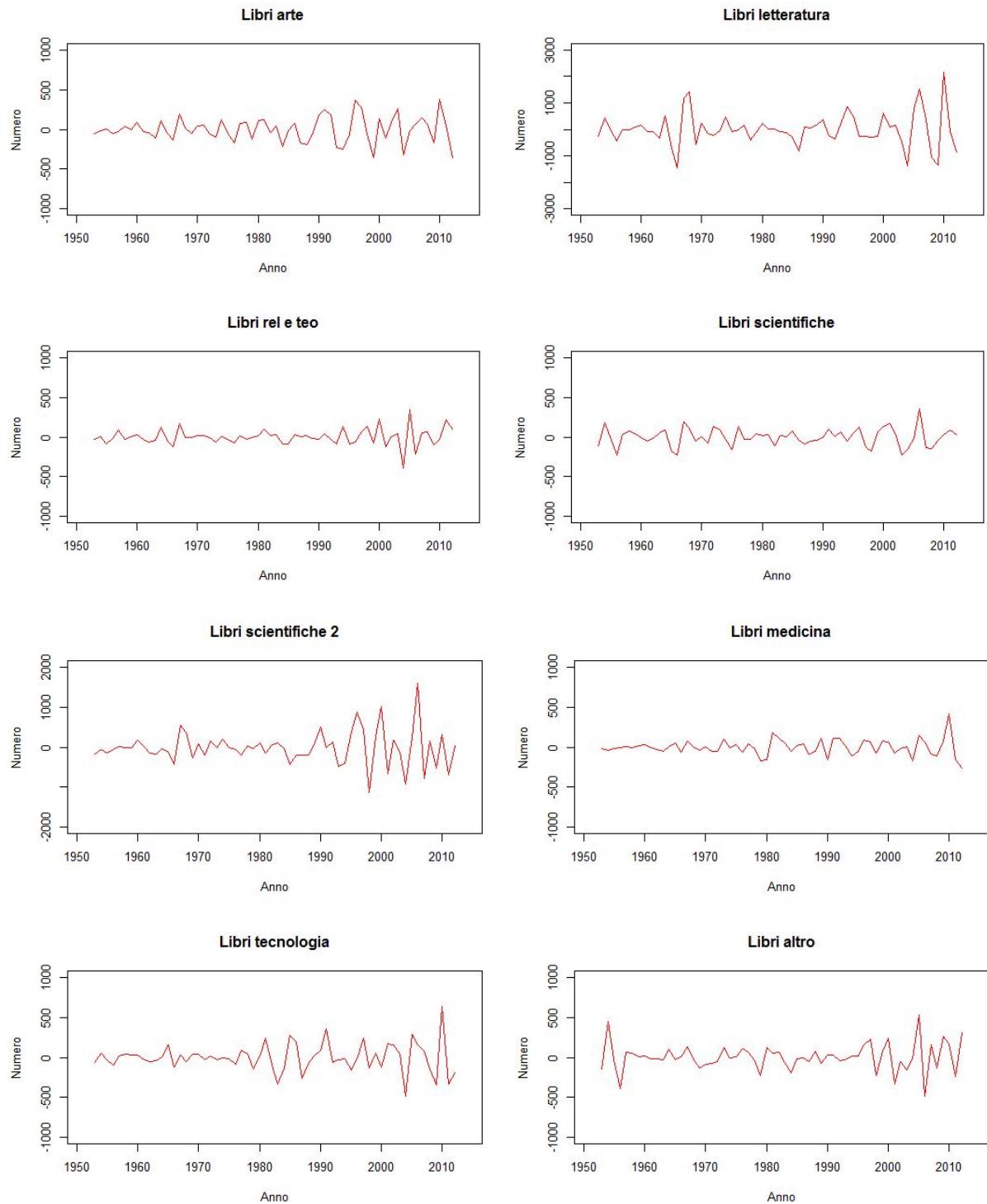


Si procede dunque, sulla base dei trend stimati, a calcolare e tracciare l'andamento delle serie attraverso la detrendizzazione.

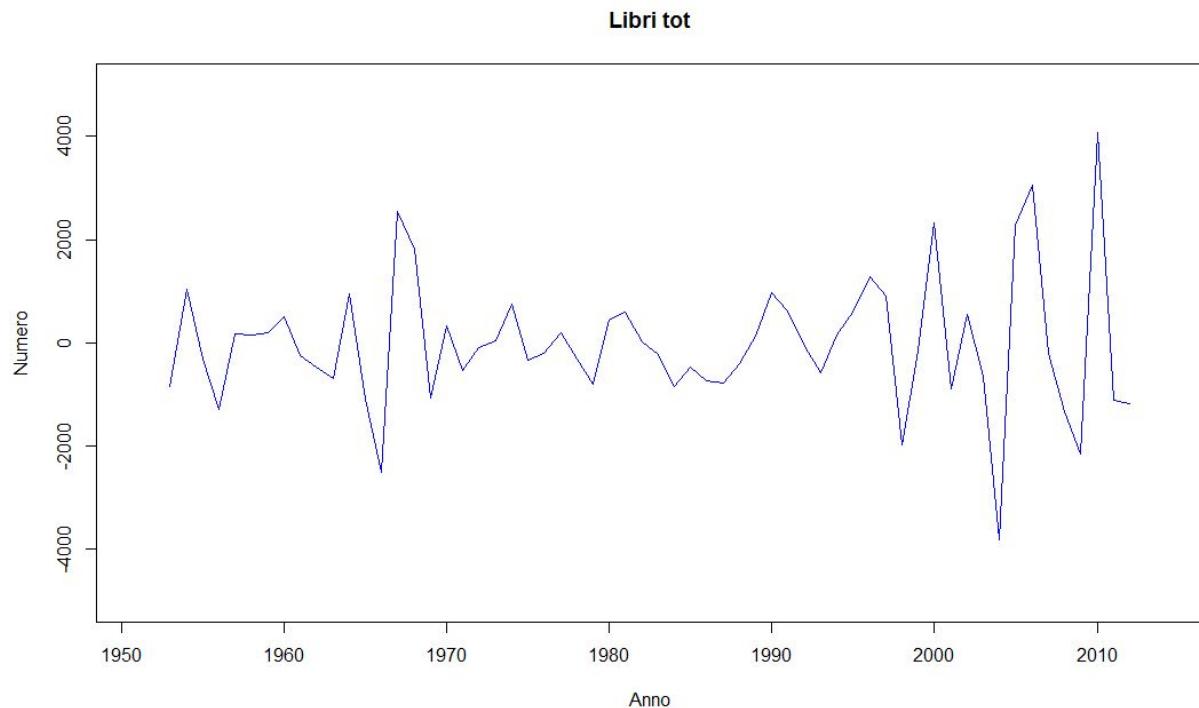
```

> detren_arte<-arti-trend_arte
> detren_lett<-letteratura-trend_letteratura
> detren_rel<-religione_e_teologia-trend_religioseteologia
> detren_sc1<-scientifiche-trend_scientifiche
> detren_sc2<-scientifiche2-trend_scientifiche2
> detren_med<-medici-trend_medicina
> detren_tech<-tecnologia-trend_tecnologia
> detren_altro<-altro-trend_altro
> par(mfrow=c(2,2))
> matplot(x, detren_arte, type = "l", ylim=c(-1000,1000), ylab = "Numero", xlab = "Anno", main="Libri arte", col="red")
> matplot(x, detren_lett, type = "l", ylim=c(-3000,3000), ylab = "Numero", xlab = "Anno", main="Libri letteratura", col="red")
> matplot(x, detren_rel, type = "l", ylim=c(-1000,1000), ylab = "Numero", xlab = "Anno", main="Libri rel e teo", col="red")
> matplot(x, detren_sc1, type = "l", ylim=c(-1000,1000), ylab = "Numero", xlab = "Anno", main="Libri scientifiche", col="red")
> matplot(x, detren_sc2, type = "l", ylim=c(-2000,2000), ylab = "Numero", xlab = "Anno", main="Libri scientifiche 2", col="red")
> matplot(x, detren_med, type = "l", ylim=c(-1000,1000), ylab = "Numero", xlab = "Anno", main="Libri medicina", col="red")
> matplot(x, detren_tech, type = "l", ylim=c(-1000,1000), ylab = "Numero", xlab = "Anno", main="Libri tecnologia", col="red")
> matplot(x, detren_altro, type = "l", ylim=c(-1000,1000), ylab = "Numero", xlab = "Anno", main="Libri i altro", col="red")

```



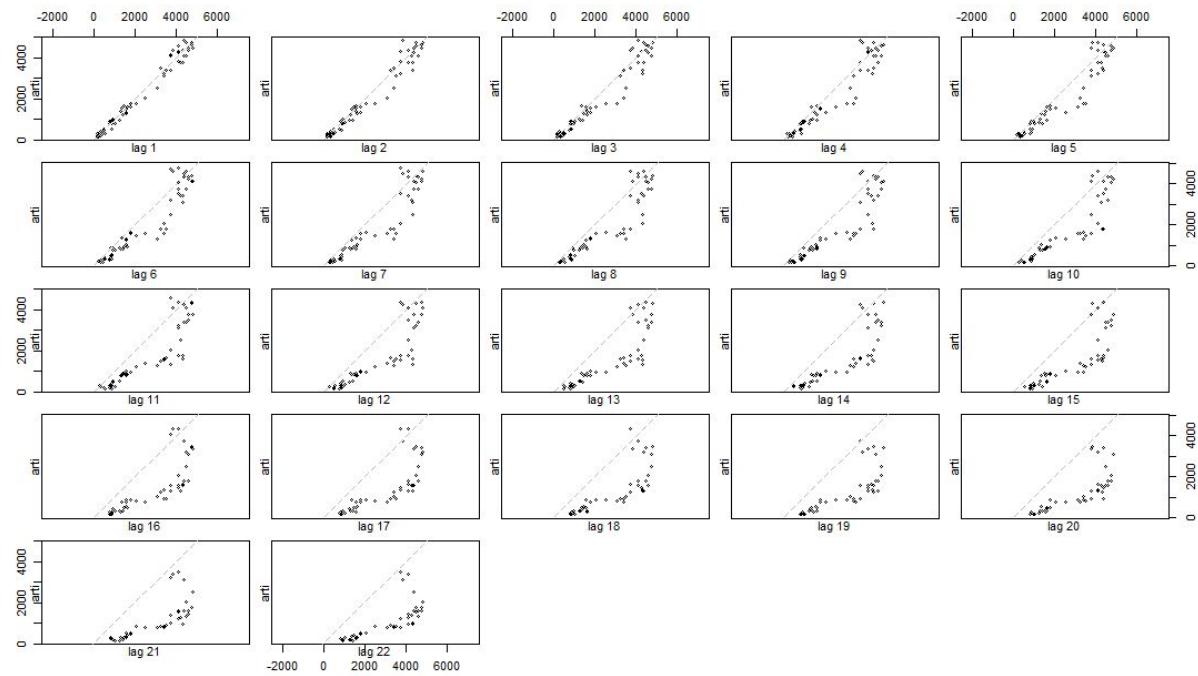
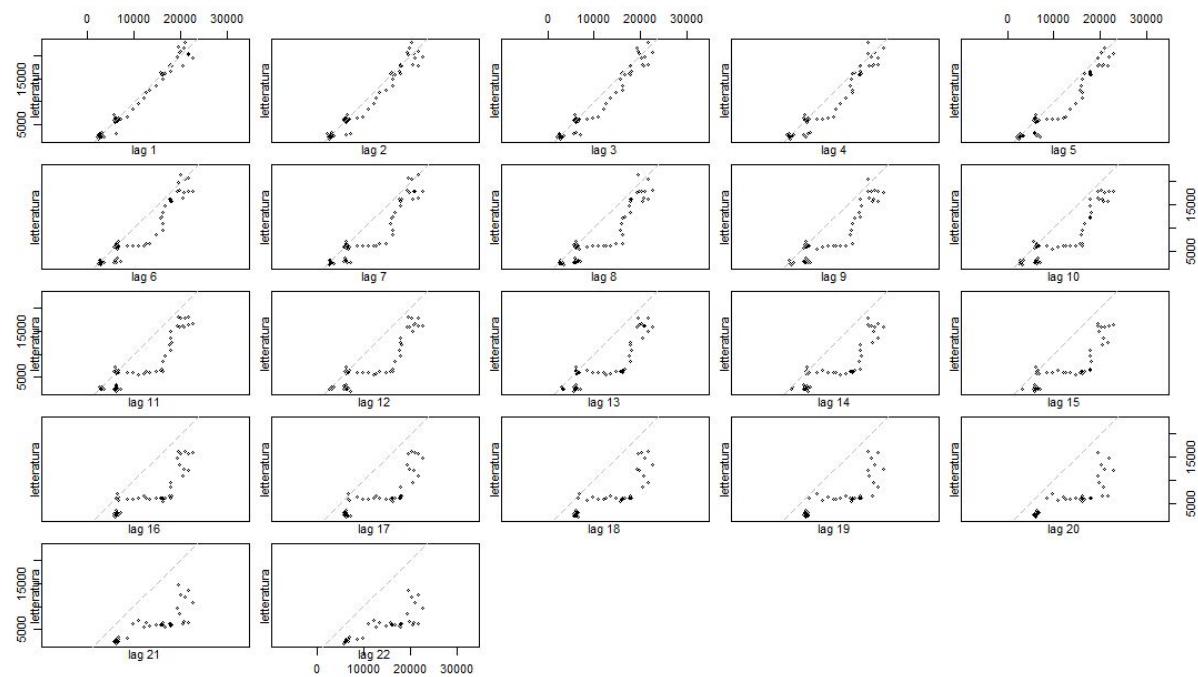
```
> detren_tot<-tot2-trend_tot
> matplot(x, detren_tot, type = "l", ylim=c(-5000,5000), ylab = "Numero", xlab = "Anno", main="Libri tot", col="blue")
```



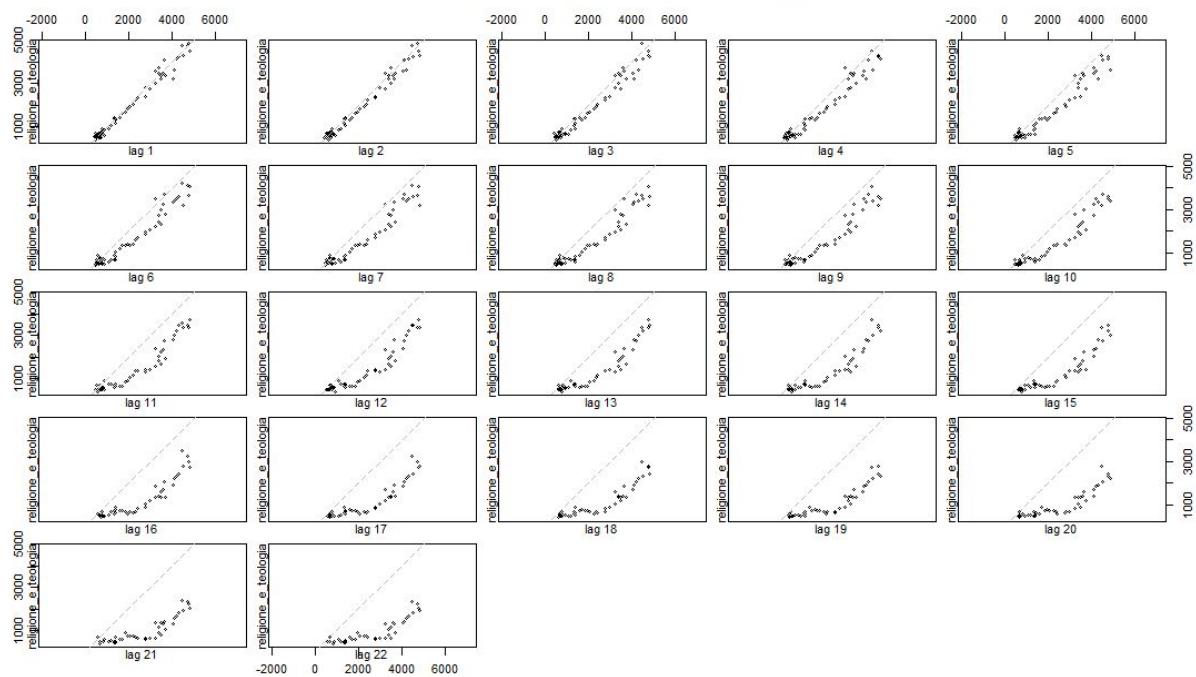
Se fosse stata usata, per la stima del trend, una media mobile di ordine 1 anziché una di ordine 2, avremmo avuto un trend più fedele alla serie originaria e quindi con più irregolarità. Di conseguenza, nella differenza, tale trend avrebbe annullato maggiormente le irregolarità della serie originaria, per cui la serie detrendizzata avrebbe presentato oscillazioni e irregolarità meno marcate.

Procediamo ora con tracciare i grafici di autodispersione.

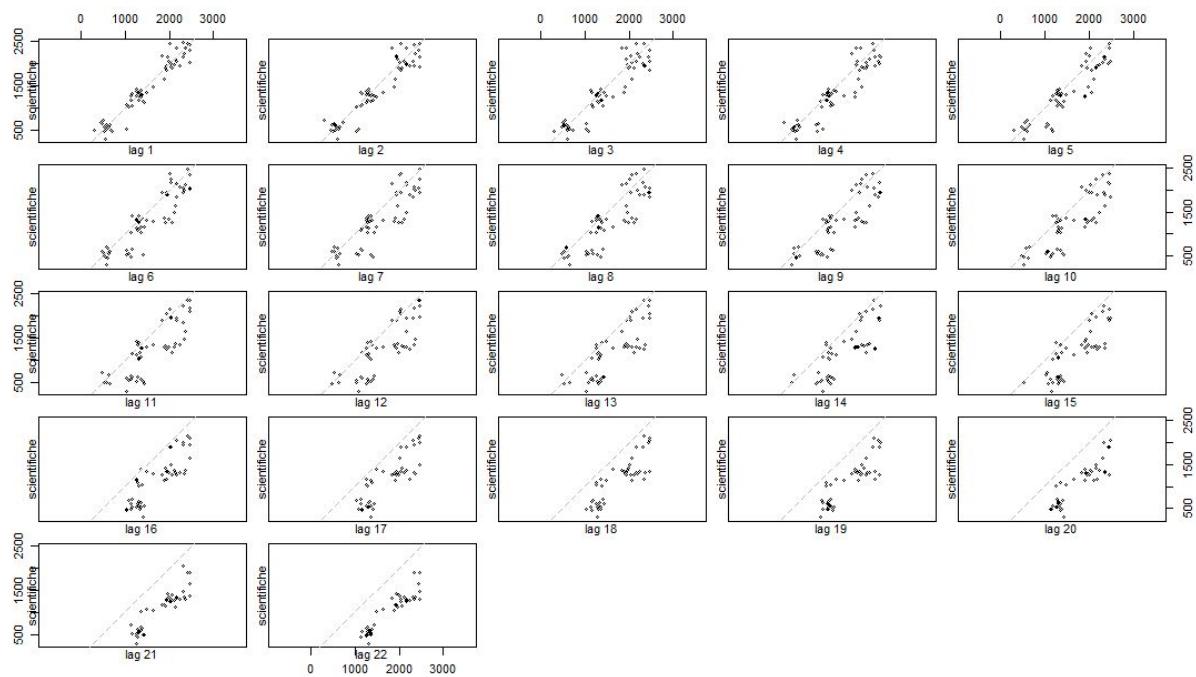
```
> lag.plot(arti, set.lags = 1:22, type = "p", do.lines=FALSE, main = "Autodispersione opere artistiche")
> lag.plot(letteratura, set.lags = 1:22, type = "p", do.lines=FALSE, main = "Autodispersione opere letterarie")
> lag.plot(religione_e_teologia, set.lags = 1:22, type = "p", do.lines=FALSE, main = "Autodispersione opere religione e teologia")
> lag.plot(scientifiche, set.lags = 1:22, type = "p", do.lines=FALSE, main = "Autodispersione opere scientifiche")
> lag.plot(scientifiche2, set.lags = 1:22, type = "p", do.lines=FALSE, main = "Autodispersione opere scientifiche2")
> lag.plot(medici, set.lags = 1:22, type = "p", do.lines=FALSE, main = "Autodispersione opere medicina")
> lag.plot(tecno, set.lags = 1:22, type = "p", do.lines=FALSE, main = "Autodispersione opere tecnologia")
> lag.plot(altro, set.lags = 1:22, type = "p", do.lines=FALSE, main = "Autodispersione opere altro")
> lag.plot(tot2, set.lags = 1:22, type = "p", do.lines=FALSE, main = "Autodispersione opere totale pubblicate")
```

**Autodispersione Opere artistiche****Autodispersione Opere letterarie**

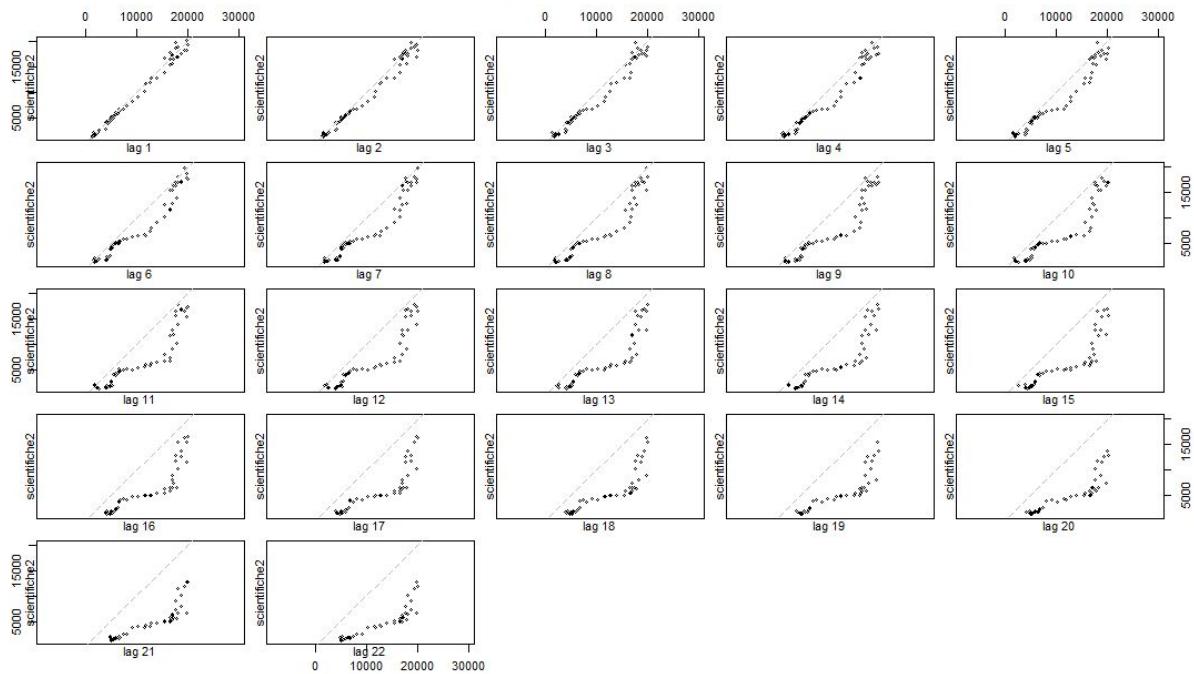
## Autodispersione Opere religione e teologia



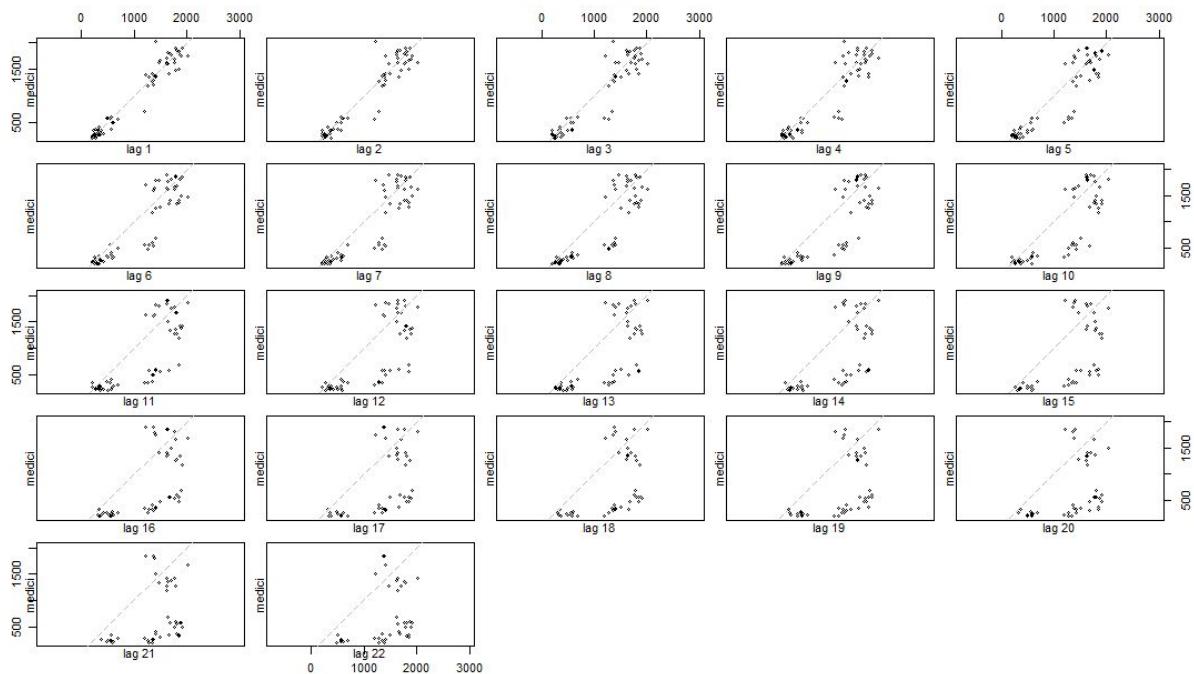
## Autodispersione Opere scientifiche



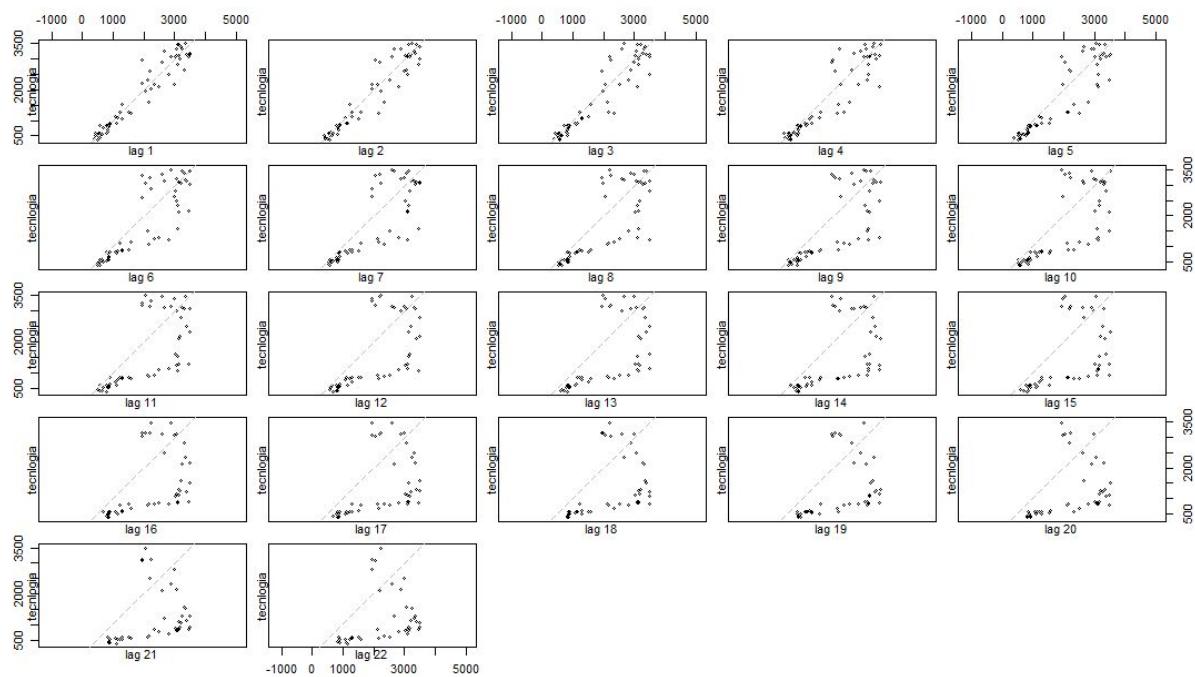
Autodispersione Opere scientifiche2



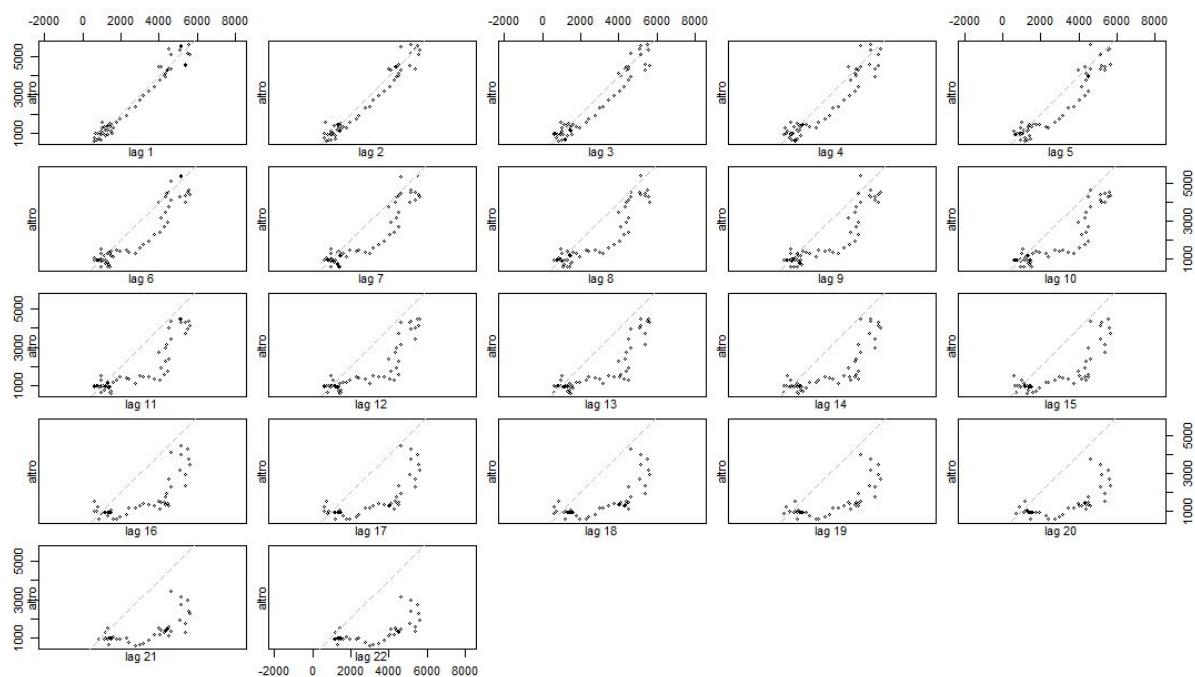
Autodispersione Opere medicina

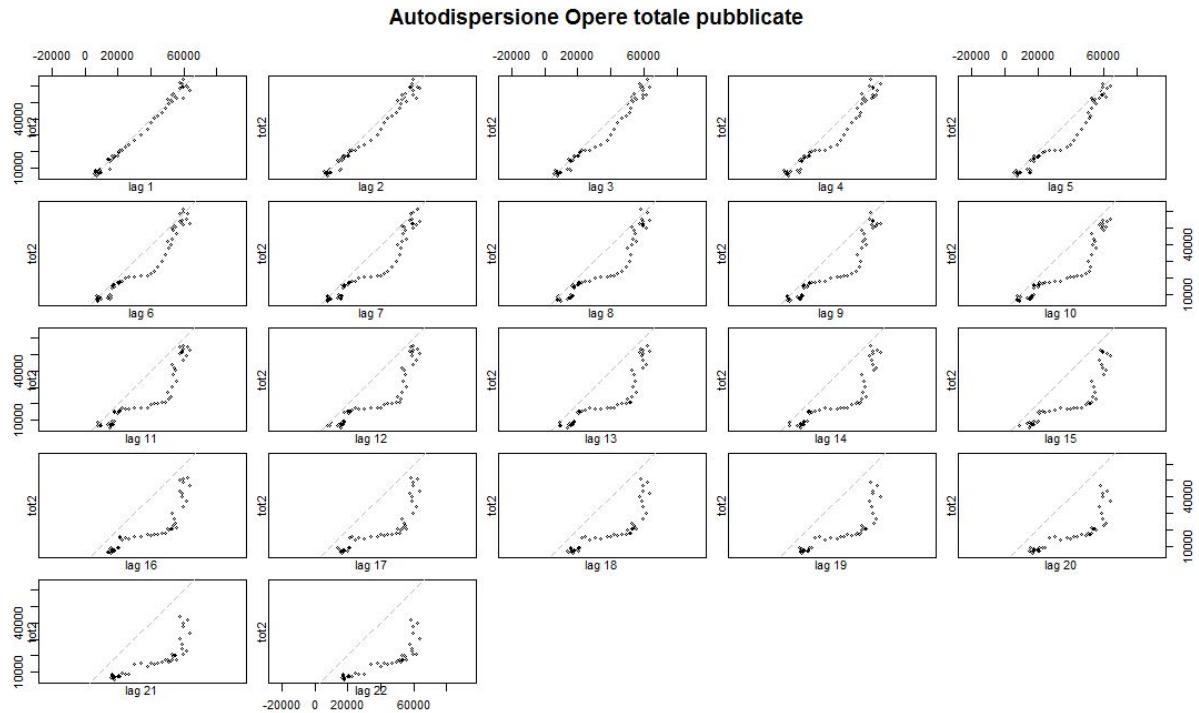


## Autodispersione Opere tecnologia



## Autodispersione Opere altro





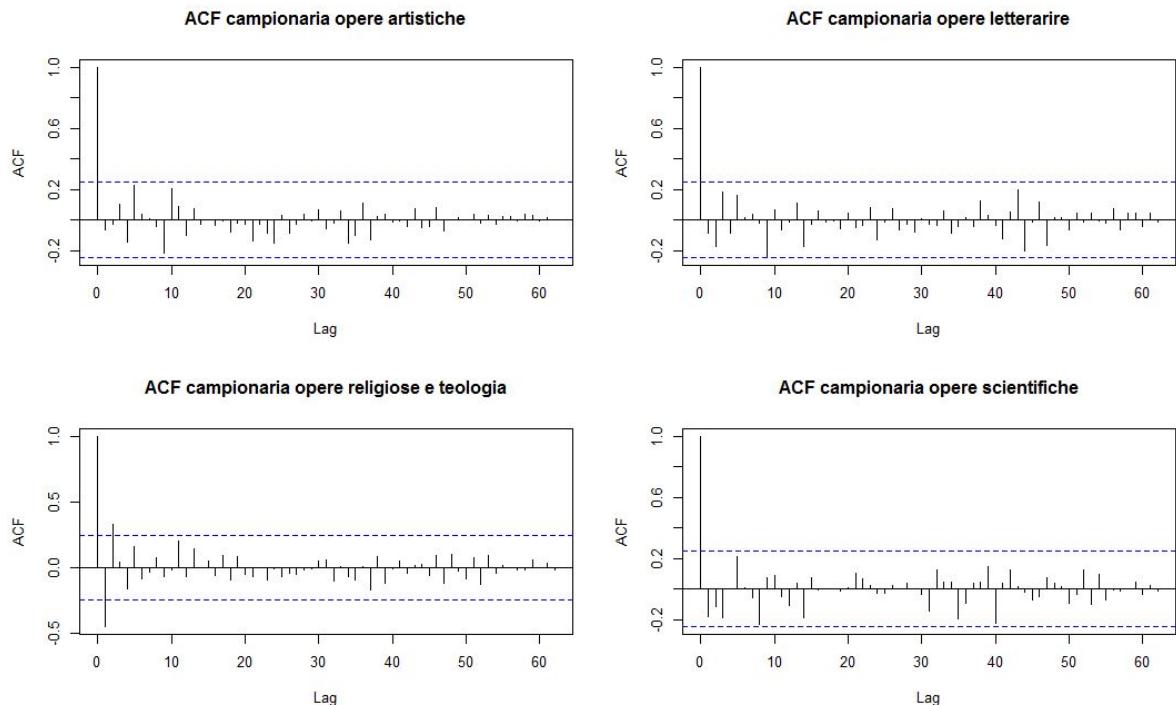
Da tali grafici, in particolare quelli inerenti alle opere librerie di tipo tecnologico, medico e scientifico, è possibile osservare che i valori assunti delle serie risultano essere abbastanza scorrelati. Questo comportamento è sempre più evidente all'aumentare della distanza tra i tempi delle osservazioni considerate, dove i valori non si dispongono lungo la retta ideale ma si disperdono sempre più caoticamente sul piano.

Come ultima analisi, si procede a detrendizzare le serie viste, mediante l'operatore differenza e ad effettuare conseguentemente il calcolo di autocorrelazione globale e parziale sulle serie risultanti dalla detrendizzazione.

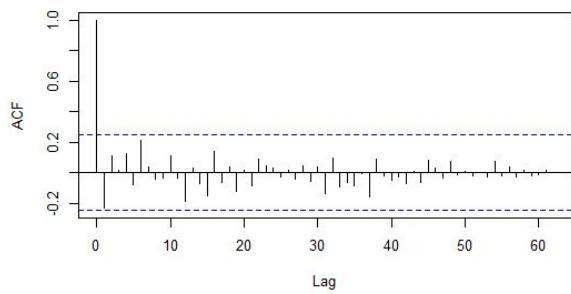
```

> diffarti<-diff(arti, lag=1, differences = 1)
> difflet<-diff(letteratura, lag=1, differences = 1)
> diffrel<-diff(religione_e_teologia, lag=1, differences = 1)
> diffsc1<-diff(scientifiche, lag=1, differences = 1)
> diffsc2<-diff(scientifiche2, lag=1, differences = 1)
> diffmed<-diff(medici, lag=1, differences = 1)
> difftech<-diff(tecnologia, lag=1, differences = 1)
> diffaltro<-diff(alter, lag=1, differences = 1)
> difftot<-diff(tot2, lag=1, differences = 1)
> par(mfrow=c(2,2))
> acf(diffarti, lag.max = 80, main="ACF campionaria opere artistiche")
> acf(difflet, lag.max = 80, main="ACF campionaria opere letterarie")
> acf(diffrel, lag.max = 80, main="ACF campionaria opere religiose e teologia")
> acf(diffsc1, lag.max = 80, main="ACF campionaria opere scientifiche")
> acf(diffsc2, lag.max = 80, main="ACF campionaria opere scientifiche 2")
> acf(diffmed, lag.max = 80, main="ACF campionaria opere mediche")
> acf(difftech, lag.max = 80, main="ACF campionaria opere tecnologiche")
> acf(diffaltro, lag.max = 80, main="ACF campionaria opere altre")
> par(mfrow=c(1,1))
> acf(difftot, lag.max = 80, main="ACF campionaria opere librerie totale")
> par(mfrow=c(2,2))
> pacf(diffarti, lag.max = 80, main="PACF campionaria opere artistiche")
> pacf(difflet, lag.max = 80, main="PACF campionaria opere letterarie")
> pacf(diffrel, lag.max = 80, main="PACF campionaria opere religiose e teologia")
> pacf(diffsc1, lag.max = 80, main="PACF campionaria opere scientifiche")
> pacf(diffsc2, lag.max = 80, main="PACF campionaria opere scientifiche 2")
> pacf(diffmed, lag.max = 80, main="PACF campionaria opere mediche")
> pacf(difftech, lag.max = 80, main="PACF campionaria opere tecnologiche")
> pacf(diffaltro, lag.max = 80, main="PACF campionaria opere altre")
> par(mfrow=c(1,1))
> pacf(difftot, lag.max = 80, main="PACF campionaria opere librerie totale")

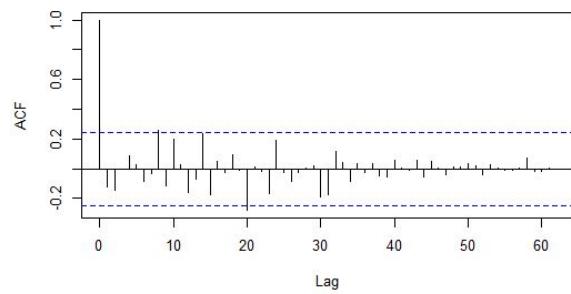
```



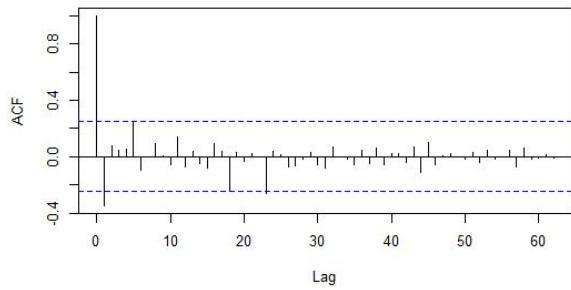
ACF campionaria opere scientifiche 2



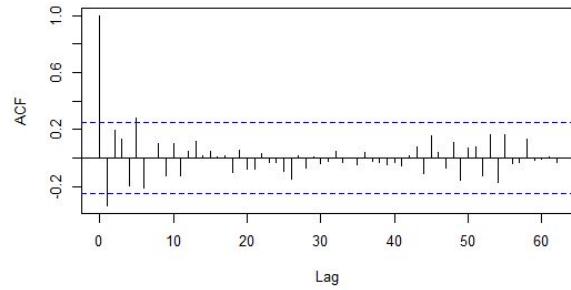
ACF campionaria opere mediche



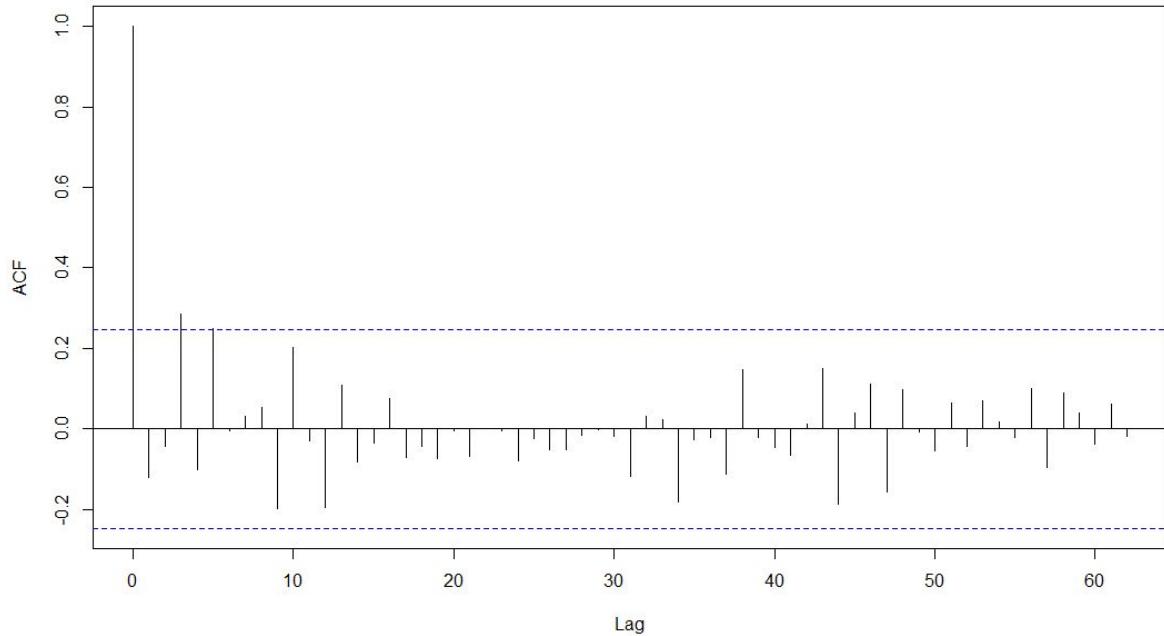
ACF campionaria opere tecnologiche

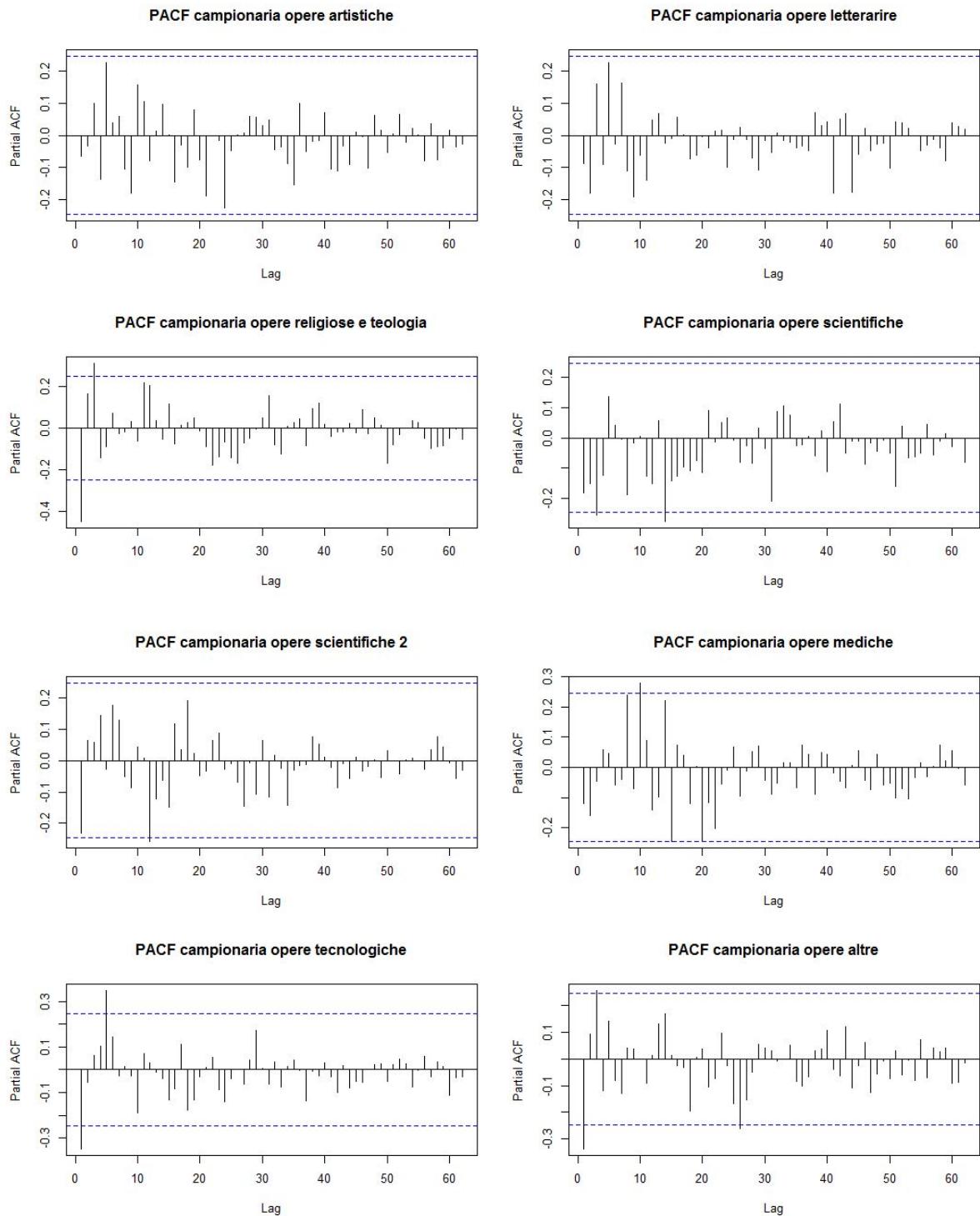


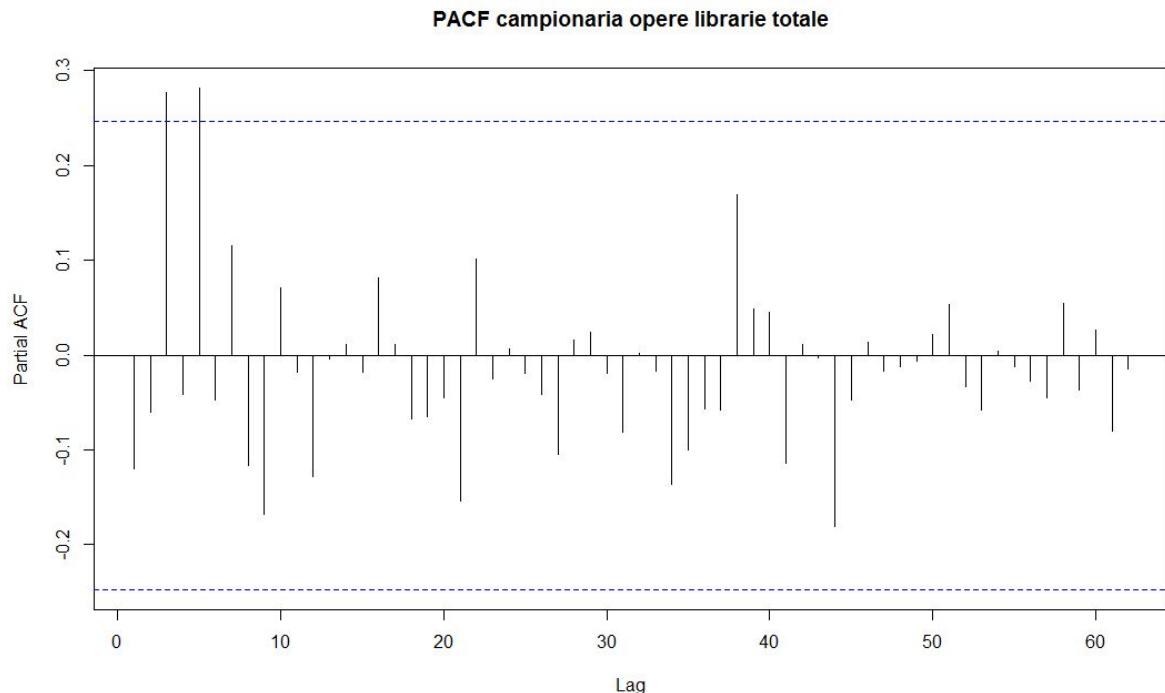
ACF campionaria opere altre



ACF campionaria opere librerie totale







Confrontando questi ultimi grafici con i corrispettivi, fatti sulla serie iniziale, il comportamento, pur non ricalcando a pieno quello di un processo stazionario noto, come ad esempio il WN, vi si potrebbe avvicinare. Infatti, i valori rimangono grosso modo contenuti entro le fasce di tolleranza. Ricordiamo comunque che ottenere, tramite simulazione o a partire da un campione di dati, un processo WN perfetto rimane pressoché impossibile. Per confermare questa somiglianza si osservano i risultati dei test di Ljung sulle serie detrendizzate.

```
> Box.test(detren_arte, type = "Ljung", lag=50)
  Box-Ljung test

data: detren_arte
X-squared = 51.65, df = 50, p-value = 0.4091

> Box.test(diffarti, type = "Ljung", lag=50)
  Box-Ljung test

data: diffarti
X-squared = 38.229, df = 50, p-value = 0.8881

> Box.test(detren_lett, type = "Ljung", lag=50)
  Box-Ljung test

data: detren_lett
X-squared = 87.617, df = 50, p-value = 0.0007977

> Box.test(difflet, type = "Ljung", lag=50)
  Box-Ljung test

data: difflet
X-squared = 59.588, df = 50, p-value = 0.1662

> Box.test(detren_rel, type = "Ljung", lag=50)
  Box-Ljung test

data: detren_rel
X-squared = 58.18, df = 50, p-value = 0.1996

> Box.test(diffrel, type = "Ljung", lag=50)
  Box-Ljung test

data: diffrel
X-squared = 60.709, df = 50, p-value = 0.1427

> Box.test(detren_sc1, type = "Ljung", lag=50)
  Box-Ljung test

data: detren_sc1
X-squared = 116.97, df = 50, p-value = 2.744e-07

> Box.test(diffsc1, type = "Ljung", lag=50)
  Box-Ljung test

data: diffsc1
X-squared = 55.086, df = 50, p-value = 0.2883
```

```
> Box.test(detren_sc2, type = "Ljung", lag=50)
  Box-Ljung test

data: detren_sc2
X-squared = 42.235, df = 50, p-value = 0.7743

> Box.test(diffsc2, type = "Ljung", lag=50)
  Box-Ljung test

data: diffsc2
X-squared = 39.966, df = 50, p-value = 0.8442

> Box.test(detren_med, type = "Ljung", lag=50)
  Box-Ljung test

data: detren_med
X-squared = 84.25, df = 50, p-value = 0.001748

> Box.test(diffmed, type = "Ljung", lag=50)
  Box-Ljung test

data: diffmed
X-squared = 54.915, df = 50, p-value = 0.2937

> Box.test(detren_tech, type = "Ljung", lag=50)
  Box-Ljung test

data: detren_tech
X-squared = 62.622, df = 50, p-value = 0.1085
```

```

> Box.test(difftech, type = "Ljung", lag=50)
  Box-Ljung test

data: difftech
X-squared = 43.998, df = 50, p-value = 0.7118

> Box.test(detren_altro, type = "Ljung", lag=50)
  Box-Ljung test

data: detren_altro
X-squared = 73.153, df = 50, p-value = 0.01805

> Box.test(diffaltro, type = "Ljung", lag=50)
  Box-Ljung test

data: diffaltro
X-squared = 60.761, df = 50, p-value = 0.1416

> Box.test(detren_tot, type = "Ljung", lag=50)
  Box-Ljung test

data: detren_tot
X-squared = 83.036, df = 50, p-value = 0.0023

> Box.test(difftot, type = "Ljung", lag=50)
  Box-Ljung test

data: difftot
X-squared = 65.836, df = 50, p-value = 0.06591

```

Dai test sopra riportati possiamo concludere che per le serie riguardanti opere librerie di tipo artistico, scientifico2 e tecnologico detrendizzati in particolare mediante l'operatore differenza la somiglianza ad un processo stazionario WN è rilevante. Per le restanti serie non è possibile affermare lo stesso.

## Stima delle emissioni inquinanti in atmosfera dal 1980-2010

Nella seguente sezione si andrà ad analizzare la serie storica riguardante le emissioni inquinanti in Italia e quindi si analizzeranno tutti quelle sostanze che incidono in maniera negativa nell'ambiente.

Tali sostanze sono:

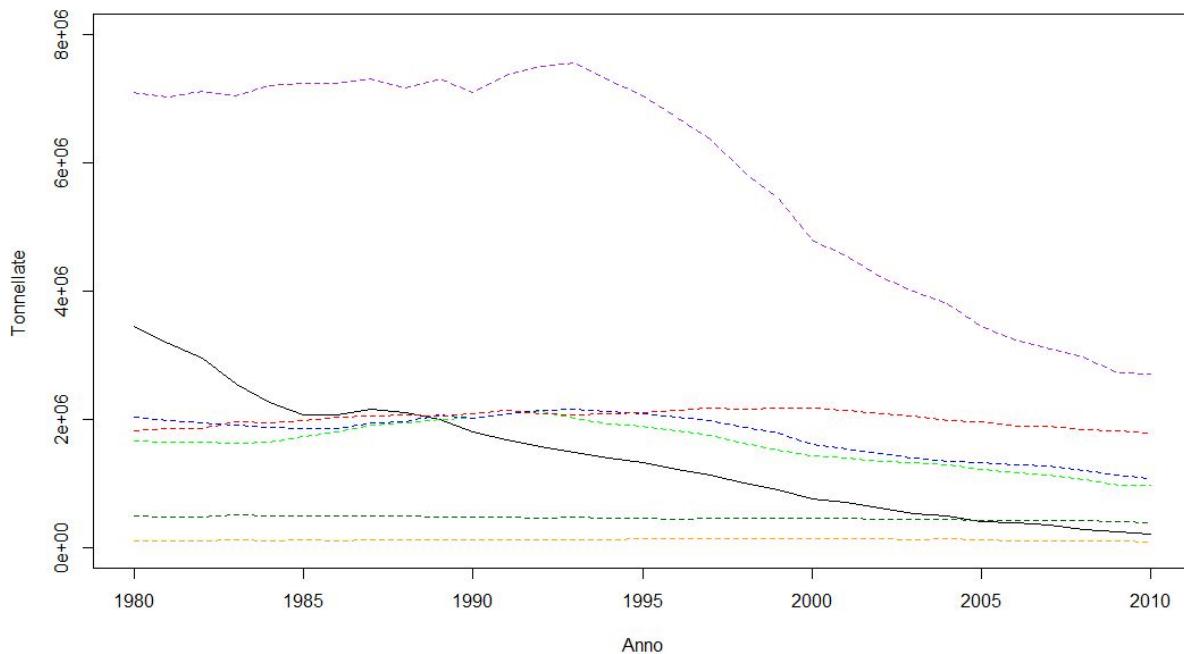
- Ossido di zolfo
- Ossido di azoto
- Composti organici volatili non metanici
- Metano
- Monossido di carbonio
- Anidride carbonica
- Protossido di azoto
- Ammoniaca

Al fine di facilitare il lettore i nomi delle sostanze sono sostituiti usando la loro formula chimica o acronimi.

- SO<sub>x</sub>
- NO<sub>x</sub>
- COVNM
- CH<sub>4</sub>
- CO
- CO<sub>2</sub>
- N<sub>2</sub>O
- NH<sub>3</sub>

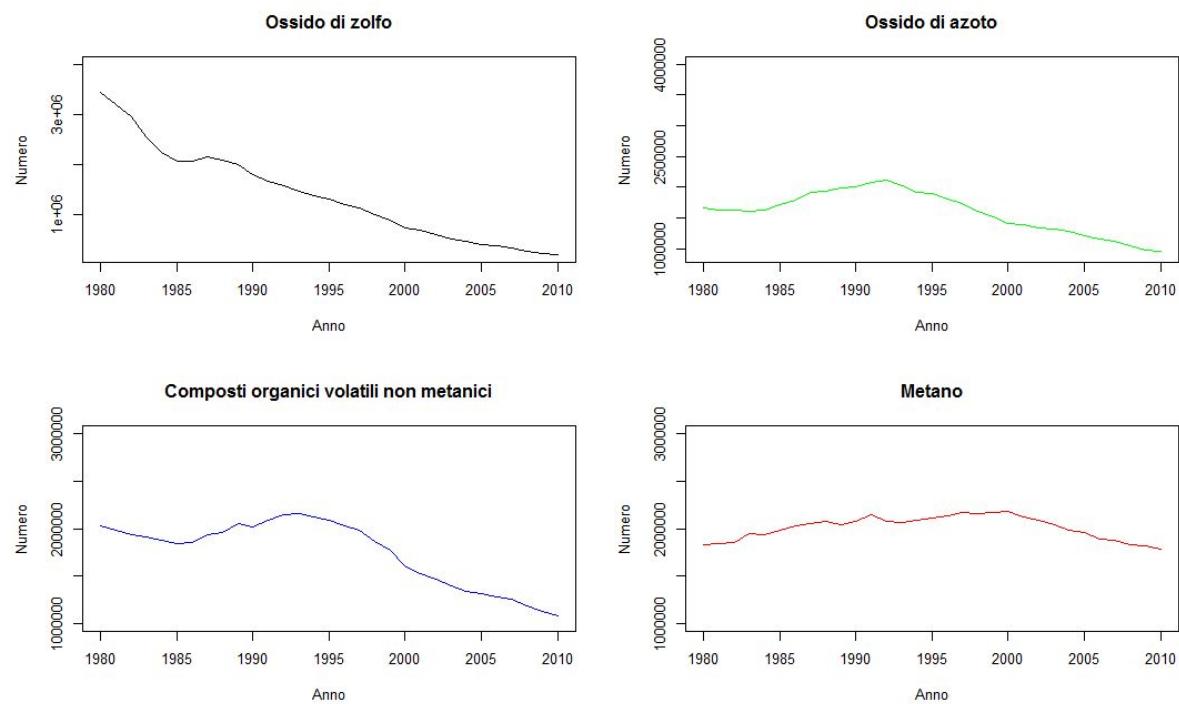
Per prima cosa portiamo in un unico grafico tutte le sostanze in modo da avere un quadro completo del loro comportamento rispetto gli anni.

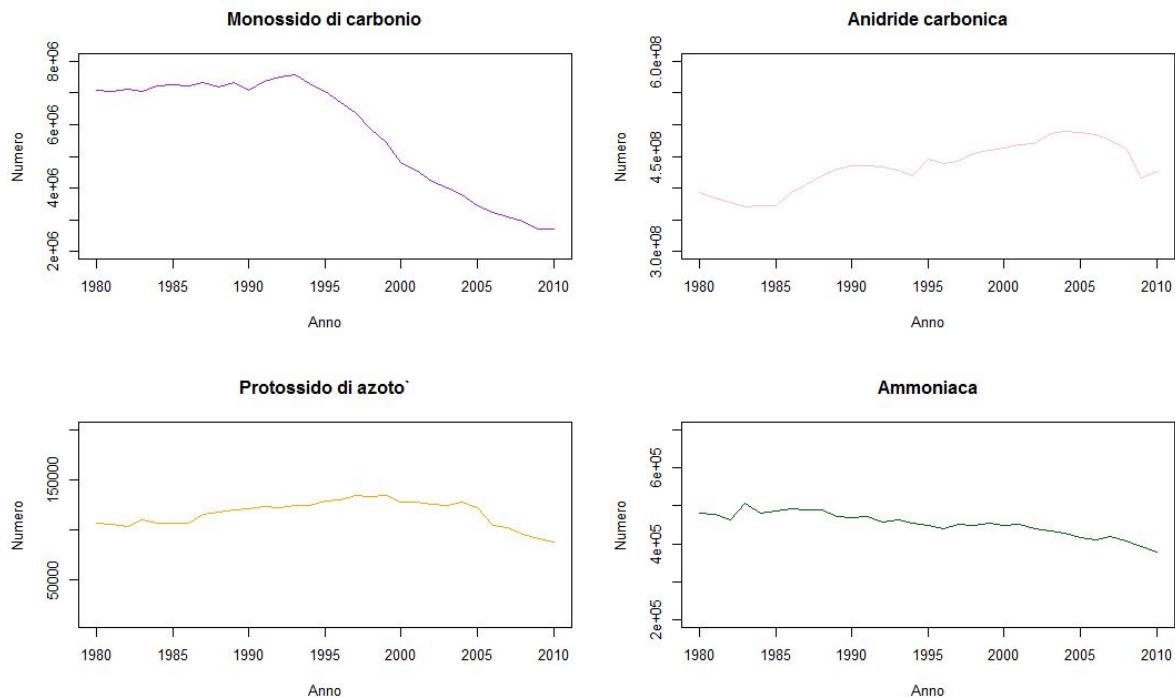
```
> x=seq(1980,2010,1)
> matplot(x,sox, type="l", ylim=c(200,8000000), ylab="Tonnellate ", xlab="Anno")
> lines(x, NOx, type="l", lty=2, col="green")
> lines(x, COVNM, type="l", lty=2, col="blue")
> lines(x,CH4 , type="l", lty=2, col="red")
> lines(x, CO, type="l", lty=2, col="purple")
> lines(x,CO2 , type="l", lty=2, col="pink")
> lines(x,N2O, type="l", lty=2, col="orange")
> lines(x, NH3, type="l", lty=2, col="darkgreen")
```



Da questa prima analisi possiamo osservare che le quantità di monossido di carbonio (in viola) e di ossido di zolfo (in nero) diminuiscono notevolmente negli anni anche se il valore del monossido di carbonio risulta essere molto elevato.

Vediamo singolarmente il loro comportamento nel tempo riportando i grafici con i colori precedentemente assegnati.





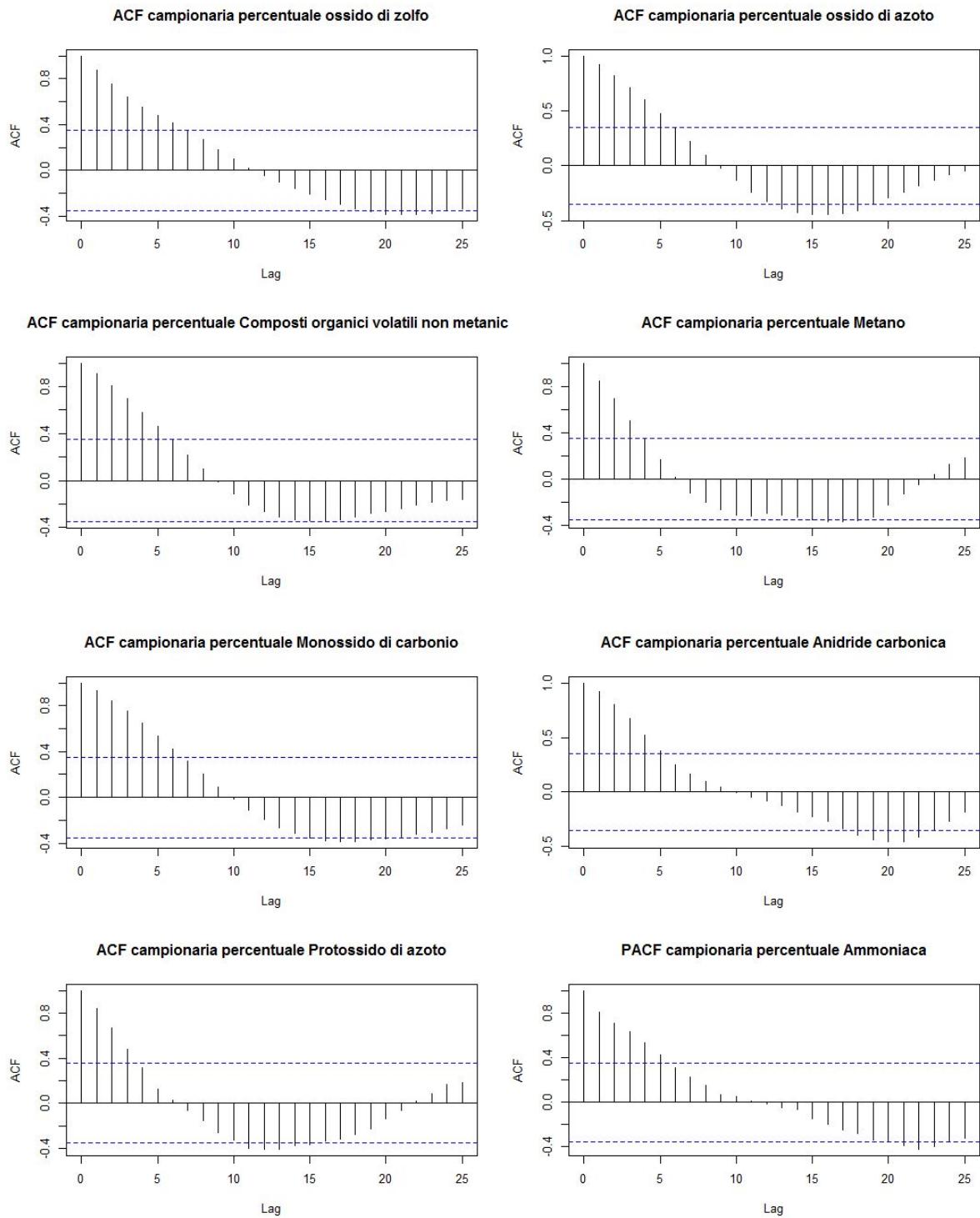
Dai grafici risulta che l'anidride carbonica ha un flusso principalmente crescente , ma fortunatamente si può osservare un calo di quasi tutte le sostanze probabilmente in seguito ai numerosi provvedimenti statali e internazionali fatti in materia dell'inquinamento.

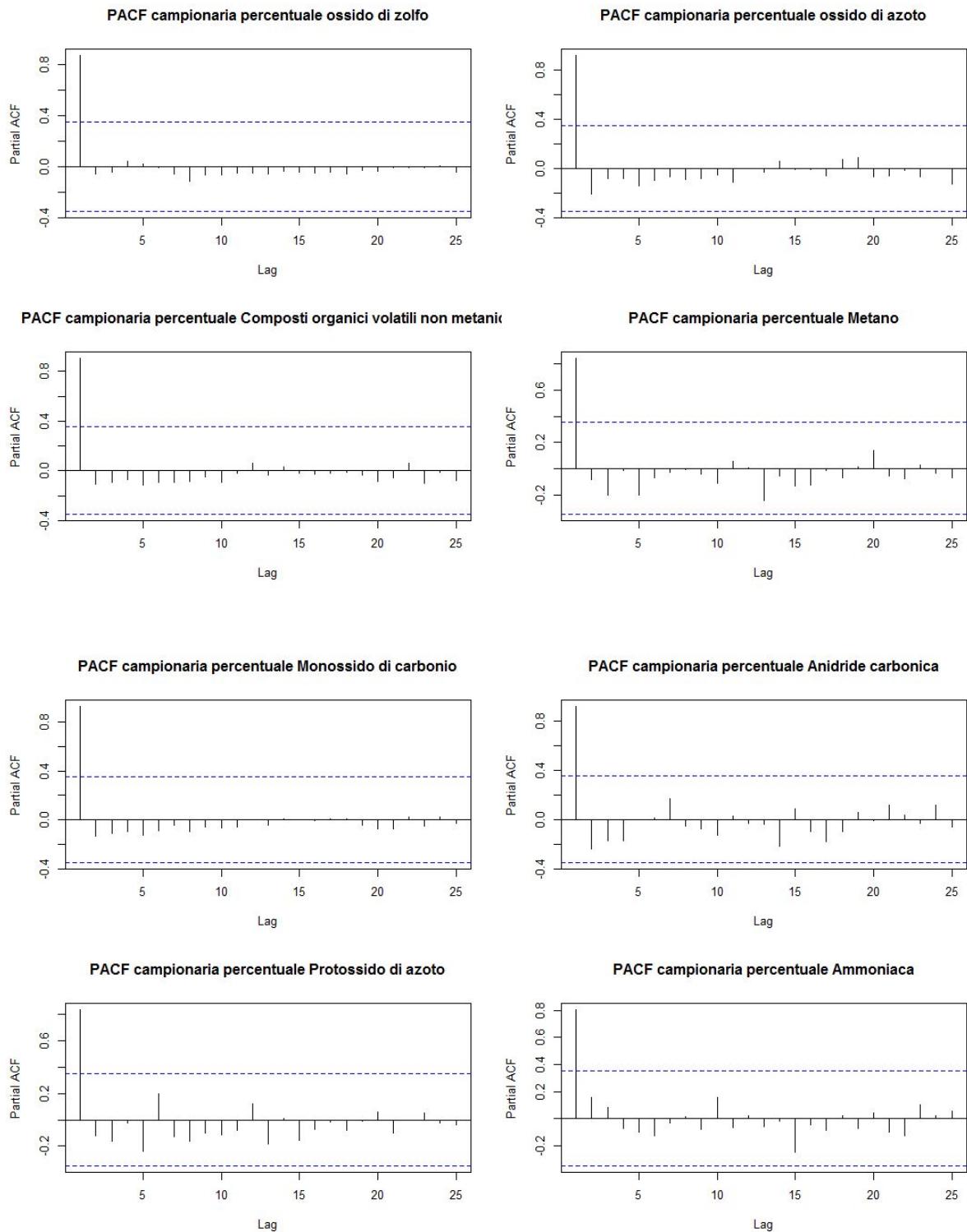
Vediamo ora se i processi risultano essere stazionari mediante i grafici di autocorrelazione globale e parziale.

```

> par(mfrow=c(2,2))
> acf(sox, lag.max =25, main="ACF campionaria percentuale ossido di zolfo")
> acf(NOX, lag.max =25, main="ACF campionaria percentuale ossido di azoto")
> acf(COVNM, lag.max =25, main="ACF campionaria percentuale Composti volatili non metanici ")
> acf(CH4, lag.max =25, main="ACF campionaria percentuale Metano")
> acf(co, lag.max =25, main="ACF campionaria percentuale Monossido di carbonio")
> acf(co2, lag.max =25, main="ACF campionaria percentuale Anidride carbonica")
> acf(N2O, lag.max =25, main="ACF campionaria percentuale Protossido di azoto")
> acf(NH3, lag.max =25, main="PACF campionaria percentuale Ammoniaca")
> par(mfrow=c(2,2))
> pacf(sox, lag.max =25, main="PACF campionaria percentuale ossido di zolfo")
> pacf(NOX, lag.max =25, main="PACF campionaria percentuale ossido di azoto")
> pacf(COVNM, lag.max =25, main="PACF campionaria percentuale composti organici volatili non metanici ")
> pacf(CH4, lag.max =25, main="PACF campionaria percentuale Metano")
> pacf(co, lag.max =25, main="PACF campionaria percentuale Monossido di carbonio")
> pacf(co2, lag.max =25, main="PACF campionaria percentuale Anidride carbonica")
> pacf(N2O, lag.max =25, main="PACF campionaria percentuale Protossido di azoto")
> pacf(NH3, lag.max =25, main="PACF campionaria percentuale Ammoniaca")

```





Da tali grafici si osserva che queste serie non si avvicinano ad un processo stazionario, inoltre è possibile dedurre che le serie considerate sono soggette a continue fluttuazioni attorno al trend principale che risulta essere quello di decrescita. Questo provoca una decrescita della funzione di autocorrelazione globale fino al raggiungimento dello 0, per poi crescere temporaneamente in modulo, ma su valori negativi.

Ulteriore conferma che i processi non siano stazionari ne tanto meno vicino ad esserlo arriva dai risultati osservabile dal test di Ljung.

```
> Box.test(SOx, type = "Ljung", lag=25)
  Box-Ljung test

  data: SOx
  X-squared = 237.47, df = 25, p-value < 2.2e-16

> Box.test(Nox, type = "Ljung", lag=25)
  Box-Ljung test

  data: Nox
  X-squared = 215.1, df = 25, p-value < 2.2e-16

> Box.test(COVNM, type = "Ljung", lag=25)
  Box-Ljung test

  data: COVNM
  X-squared = 183.49, df = 25, p-value < 2.2e-16

> Box.test(CH4, type = "Ljung", lag=25)
  Box-Ljung test

  data: CH4
  X-squared = 151.8, df = 25, p-value < 2.2e-16

> Box.test(CO, type = "Ljung", lag=25)
  Box-Ljung test

  data: CO
  X-squared = 250.85, df = 25, p-value < 2.2e-16

> Box.test(CO2, type = "Ljung", lag=25)
  Box-Ljung test

  data: CO2
  X-squared = 228.22, df = 25, p-value < 2.2e-16

> Box.test(N2O, type = "Ljung", lag=25)
  Box-Ljung test

  data: N2O
  X-squared = 144.16, df = 25, p-value < 2.2e-16

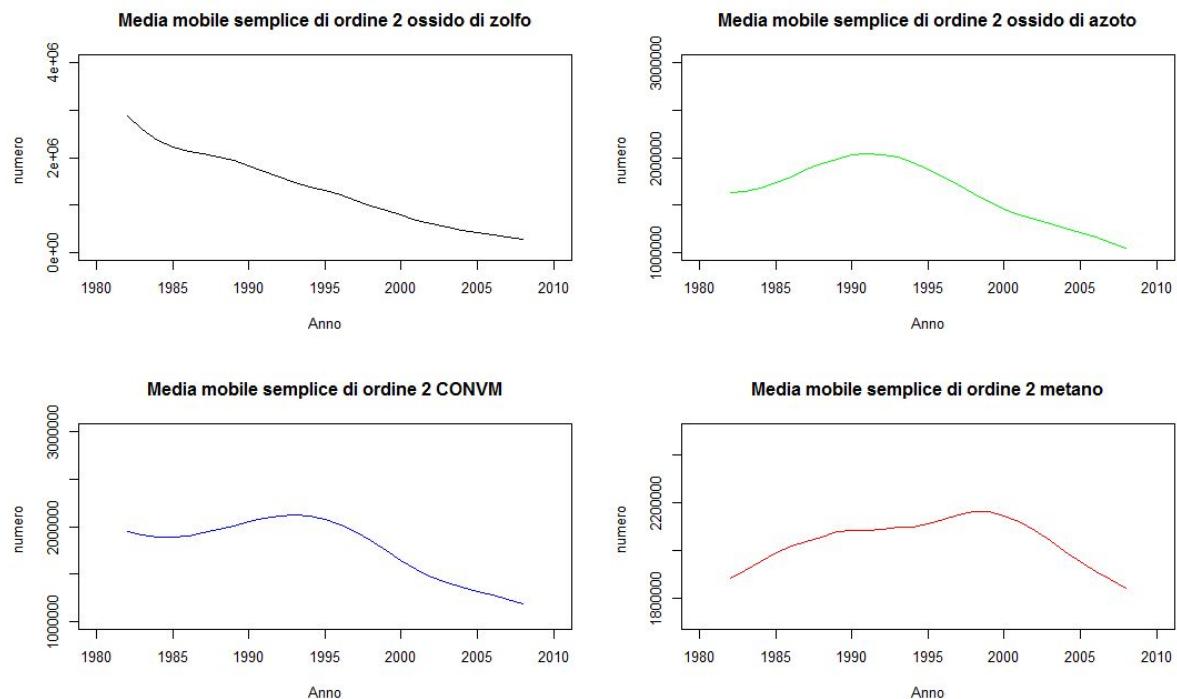
> Box.test(NH3, type = "Ljung", lag=25)
  Box-Ljung test

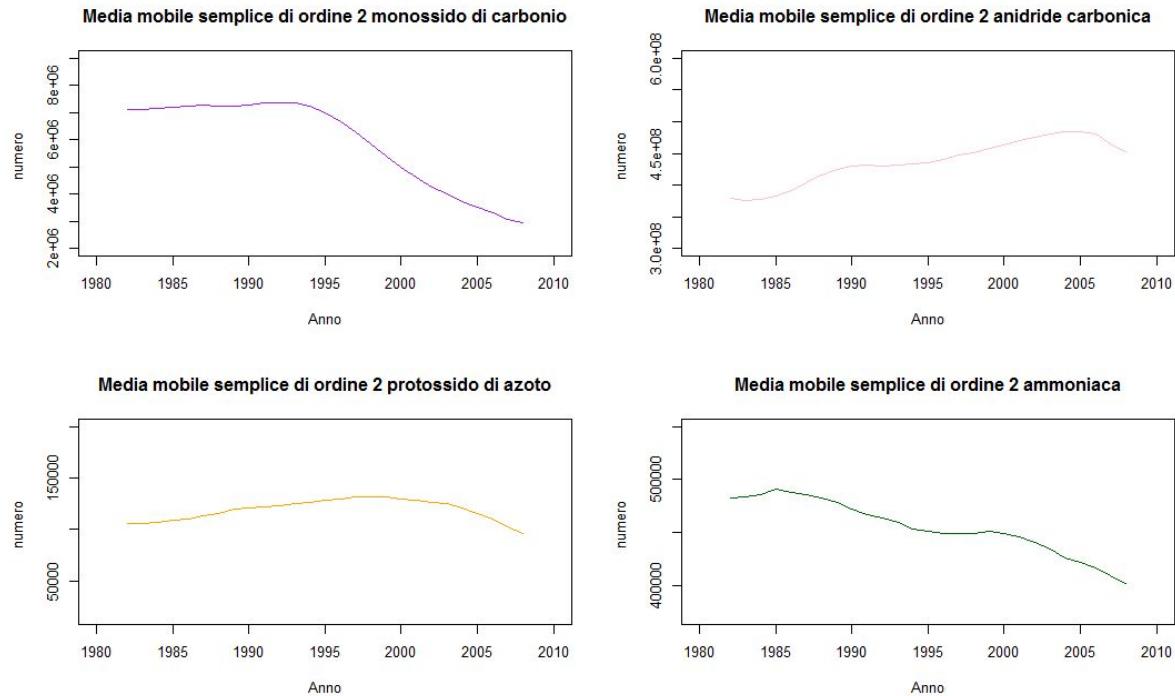
  data: NH3
  X-squared = 207.75, df = 25, p-value < 2.2e-16
```

Come atteso, tutti i valori risultano essere notevolmente bassi.

Per quanto riguarda la stima del trend, considerando che i dati a disposizione riguardano soltanto 30 anni, per evitare un fenomeno simile all'overfitting, si ritiene più opportuno effettuare tale stima usando una media mobile semplice di ordine 2, anziché di ordine 1, così da acquisire in misura minore, all'interno del trend, le possibili irregolarità presenti nella serie.

```
> alpha<-rep(1/5,5) #ordine 2
> trend_SOX<-filter(SOX, alpha, method = "convolution", sides=2, circular = FALSE)
> trend_NOX<-filter(NOX, alpha, method = "convolution", sides=2, circular = FALSE)
> trend_COVM<-filter(COVM, alpha, method = "convolution", sides=2, circular = FALSE)
> trend_CH4<-filter(CH4, alpha, method = "convolution", sides=2, circular = FALSE)
> trend_CO<-filter(CO, alpha, method = "convolution", sides=2, circular = FALSE)
> trend_CO2<-filter(CO2, alpha, method = "convolution", sides=2, circular = FALSE)
> trend_N2O<-filter(N2O, alpha, method = "convolution", sides=2, circular = FALSE)
> trend_NH3<-filter(NH3, alpha, method = "convolution", sides=2, circular = FALSE)
> par(mfrow=c(2,2))
> matplot(x,trend_SOX,col = "black", type="l", ylim=c(0,4000000), ylab="numero ", xlab="Anno", main="Media mobile semplice di ordine 2 ossido di zolfo")
> matplot(x,trend_NOX,col = "green", type="l", ylim=c(1000000,3000000), ylab="numero ", xlab="Anno", main="Media mobile semplice di ordine 2 ossido di azoto")
> matplot(x,trend_COVM,col = "blue", type="l", ylim=c(1000000,3000000), ylab="numero ", xlab="Anno", main="Media mobile semplice di ordine 2 CONVM")
> matplot(x,trend_CH4,col = "red", type="l", ylim=c(1700000,2500000), ylab="numero ", xlab="Anno", main="Media mobile semplice di ordine 2 metano")
> matplot(x,trend_CO,col = "purple", type="l", ylim=c(2000000,9000000), ylab="numero ", xlab="Anno", main="Media mobile semplice di ordine 2 monossido di carbonio")
> matplot(x,trend_CO2,col = "pink", type="l", ylim=c(300000000,600000000), ylab="numero ", xlab="Anno", main="Media mobile semplice di ordine 2 anidride carbonica")
> matplot(x,trend_N2O,col = "orange", type="l", ylim=c(15000,200000), ylab="numero ", xlab="Anno", main="Media mobile semplice di ordine 2 protossido di azoto")
> matplot(x,trend_NH3,col = "darkgreen", type="l", ylim=c(370000,550000), ylab="numero ", xlab="Anno", main="Media mobile semplice di ordine 2 ammoniaca")
```



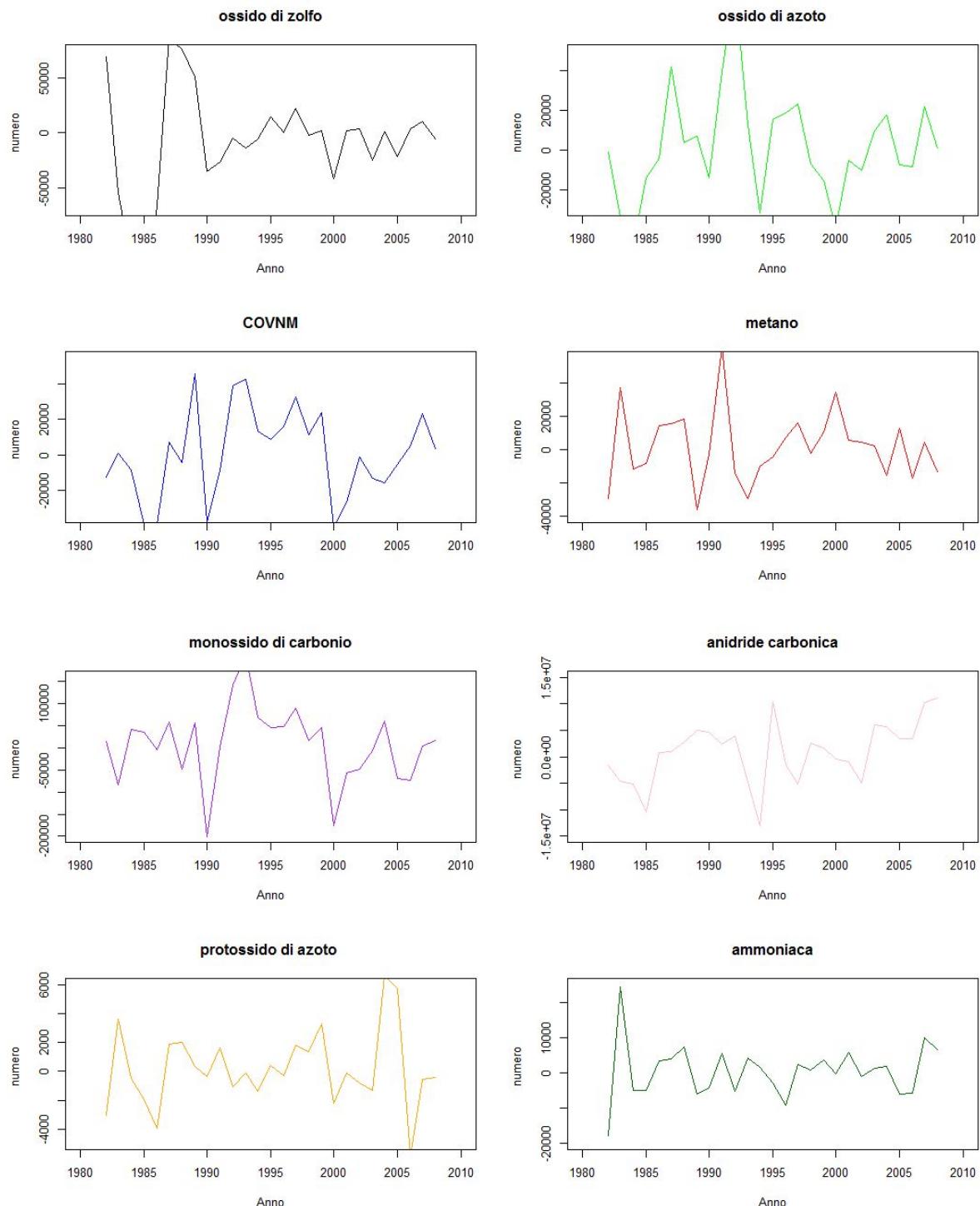


A partire dai trend generati, si elaborano le serie detrendizzate. Di seguito i comandi e i relativi grafici.

```

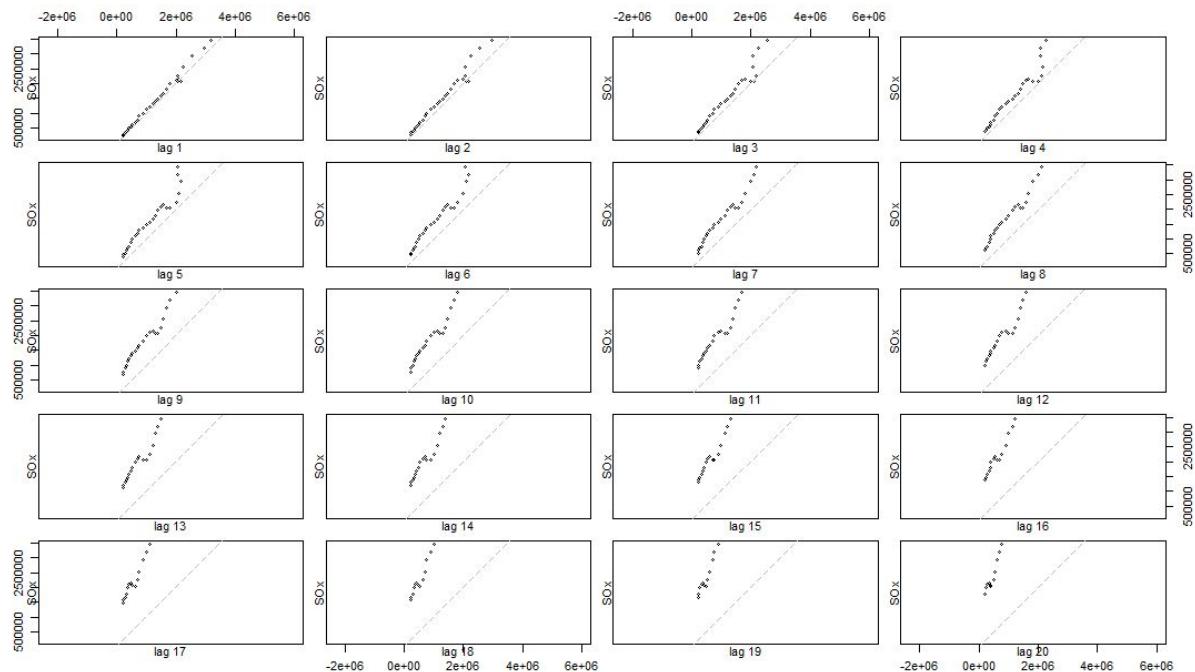
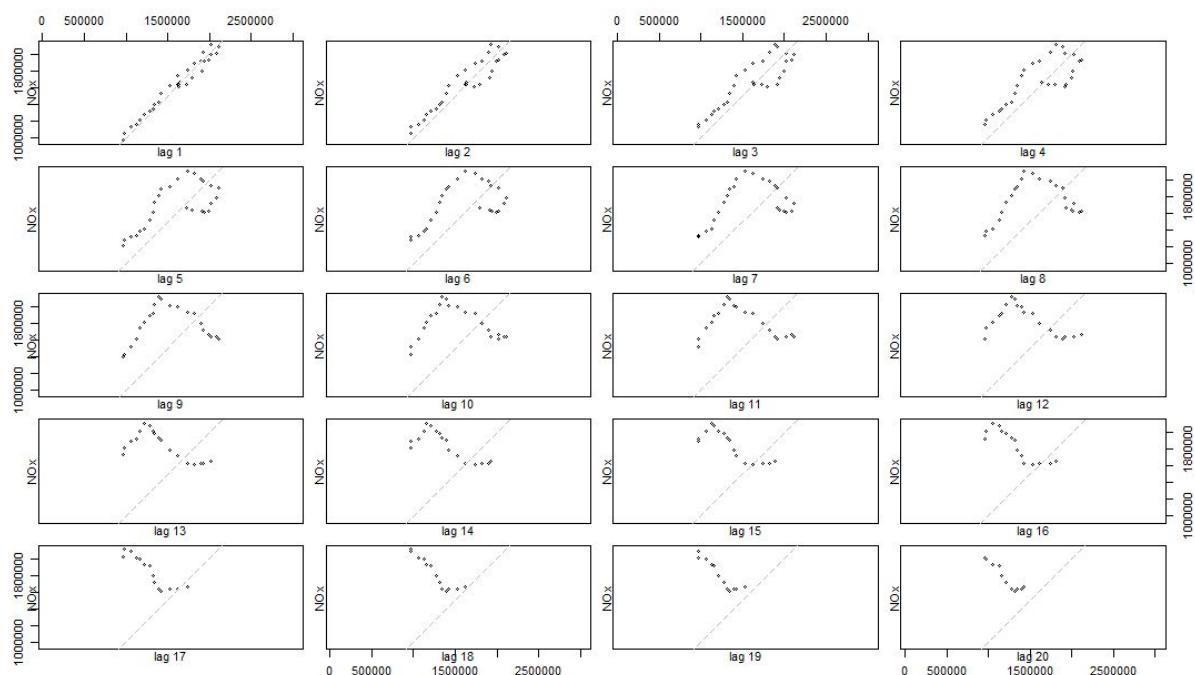
> par(mfrow=c(2,2))
> detren_SOx<-SOx-trend_so
> detren_NOx<-NOx-trend_no
> detren_COVM<-COVM-trend_COVM
> detren_CH4<-CH4-trend_CH4
> detren_CO<-CO-trend_CO
> detren_CO2<-CO2-trend_CO2
> detren_N2O<-N2O-trend_N2O
> detren_NH3<-NH3-trend_NH3
> matplot(x,detren_so, col = "black", type="l", ylim=c(-70000,75000), ylab="numero ", xlab="Anno", main="ossido di zolfo")
> matplot(x,detren_no, col = "green", type="l", ylim=c(-30000,50000), ylab="numero ", xlab="Anno", main="ossido di azoto")
> matplot(x,detren_COVM, col = "blue", type="l", ylim=c(-35000,55000), ylab="numero ", xlab="Anno", main="COVM")
> matplot(x,detren_CH4, col = "red", type="l", ylim=c(-40000,55000), ylab="numero ", xlab="Anno", main="etano")
> matplot(x,detren_CO, col = "purple", type="l", ylim=c(-200000,160000), ylab="numero ", xlab="Anno", main="monossido di carbonio")
> matplot(x,detren_CO2, col = "pink", type="l", ylim=c(-15000000,15000000), ylab="numero ", xlab="Anno", main="anidride carbonica")
> matplot(x,detren_N2O, col = "orange", type="l", ylim=c(-5000,6000), ylab="numero ", xlab="Anno", main="protossido di azoto ")
> matplot(x,detren_NH3, col = "darkgreen", type="l", ylim=c(-20000,25000), ylab="numero ", xlab="Anno", main="ammoniaca")

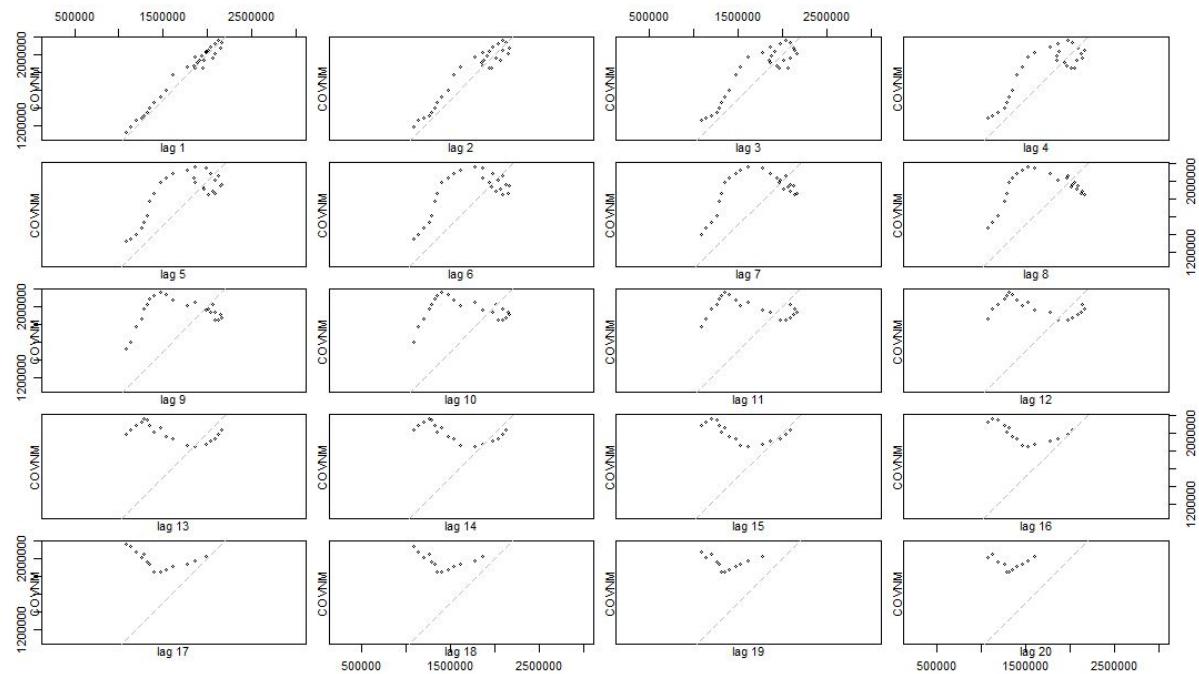
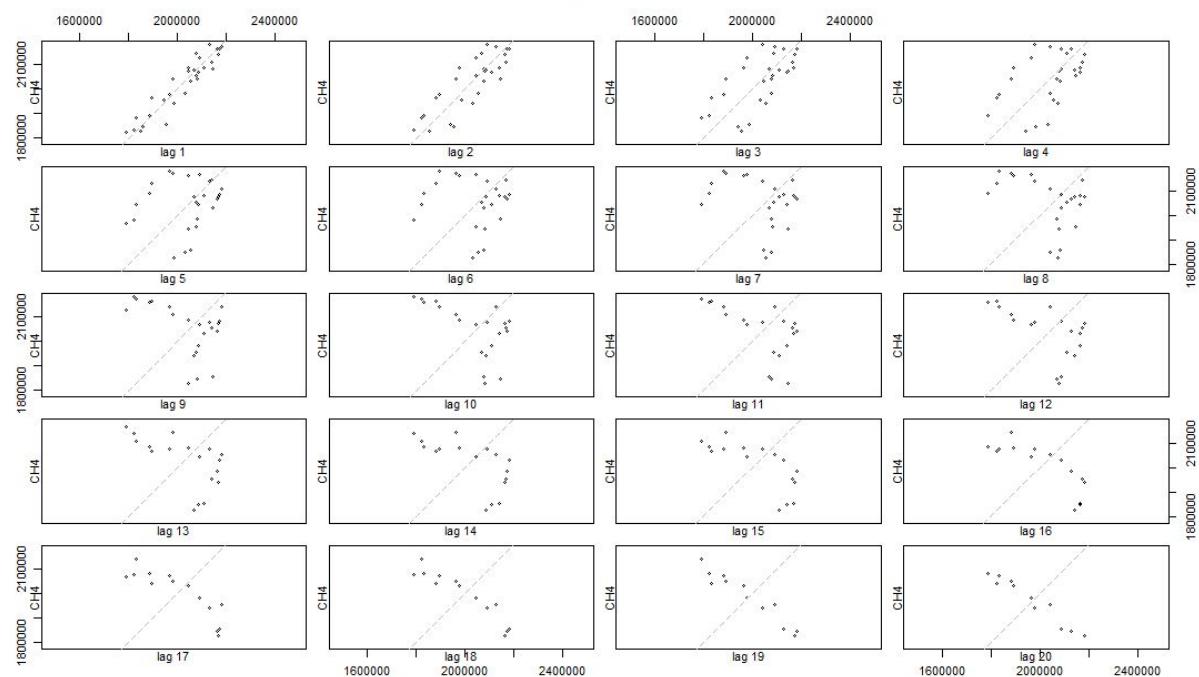
```



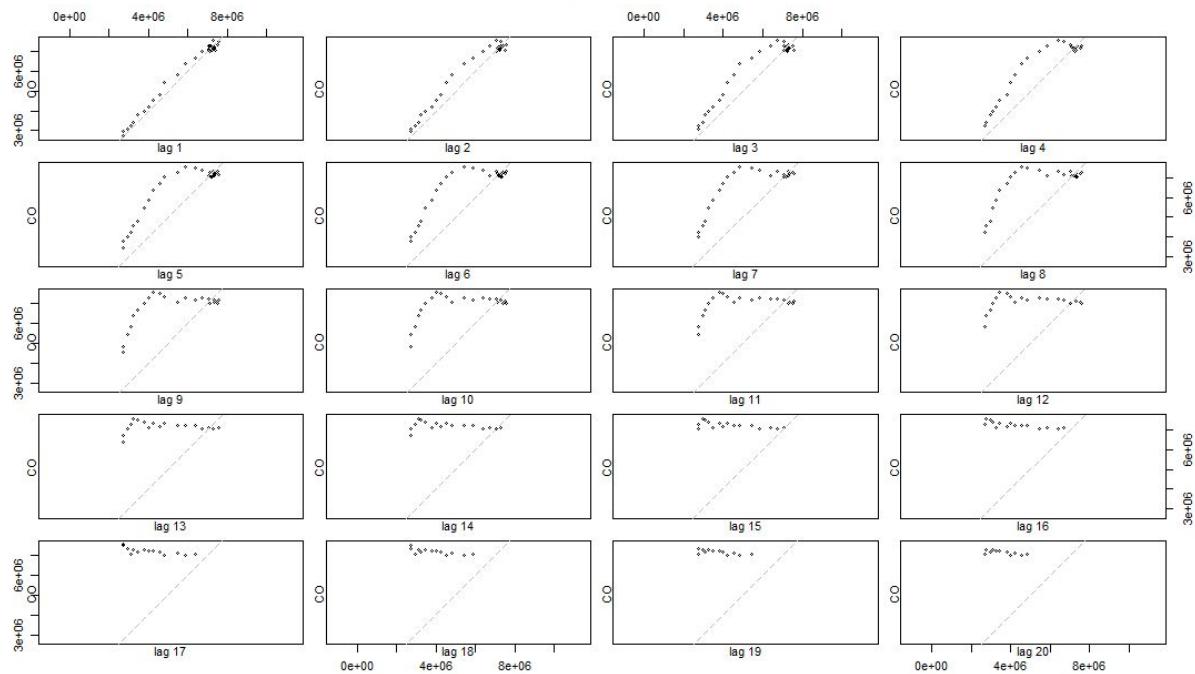
Si passa ora a tracciare ed esaminare i grafici di autodispersione delle serie.

```
> lag.plot(SOX, set.lags = 1:20, type = "p", do.lines=FALSE, main = "Autodispersione SOX" )
> lag.plot(NOX, set.lags = 1:20, type = "p", do.lines=FALSE, main = "Autodispersione NOx" )
> lag.plot(COVNM, set.lags = 1:20, type = "p", do.lines=FALSE, main = "Autodispersione COVNM" )
> lag.plot(CH4, set.lags = 1:20, type = "p", do.lines=FALSE, main = "Autodispersione CH4" )
> lag.plot(CO, set.lags = 1:20, type = "p", do.lines=FALSE, main = "Autodispersione CO" )
> lag.plot(CO2, set.lags = 1:20, type = "p", do.lines=FALSE, main = "Autodispersione CO2" )
> lag.plot(N2O, set.lags = 1:20, type = "p", do.lines=FALSE, main = "Autodispersione N2O" )
> lag.plot(NH3, set.lags = 1:20, type = "p", do.lines=FALSE, main = "Autodispersione NH3" )
```

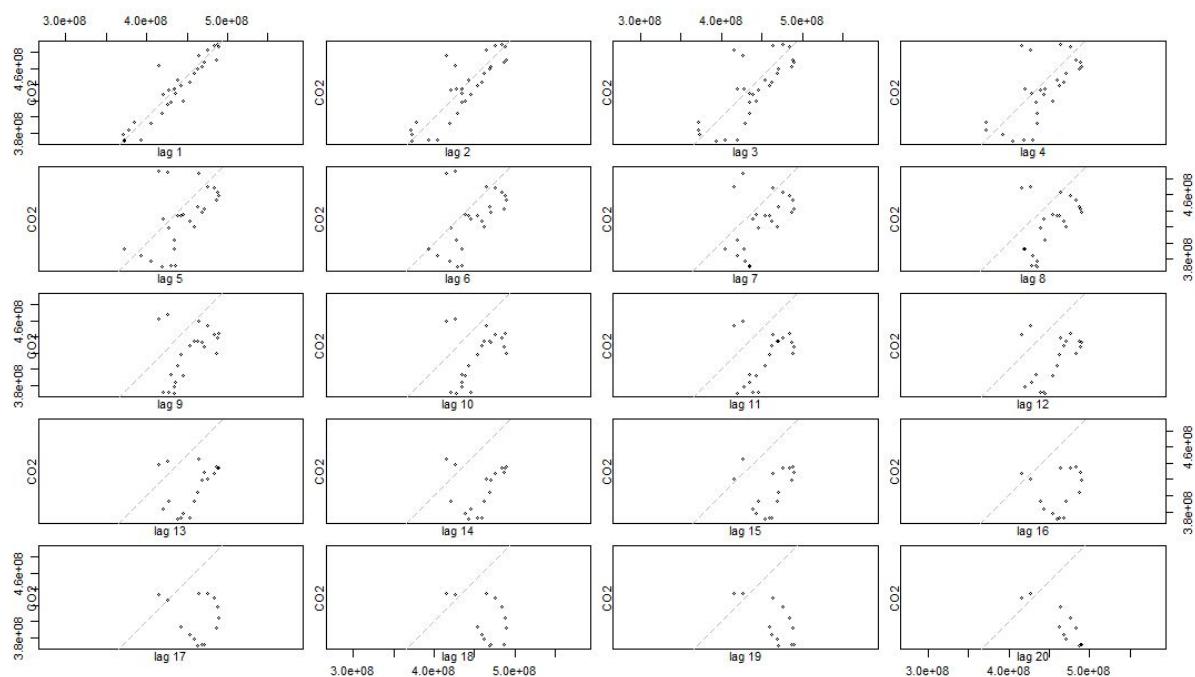
**Autodispersione SOX****Autodispersione NOx**

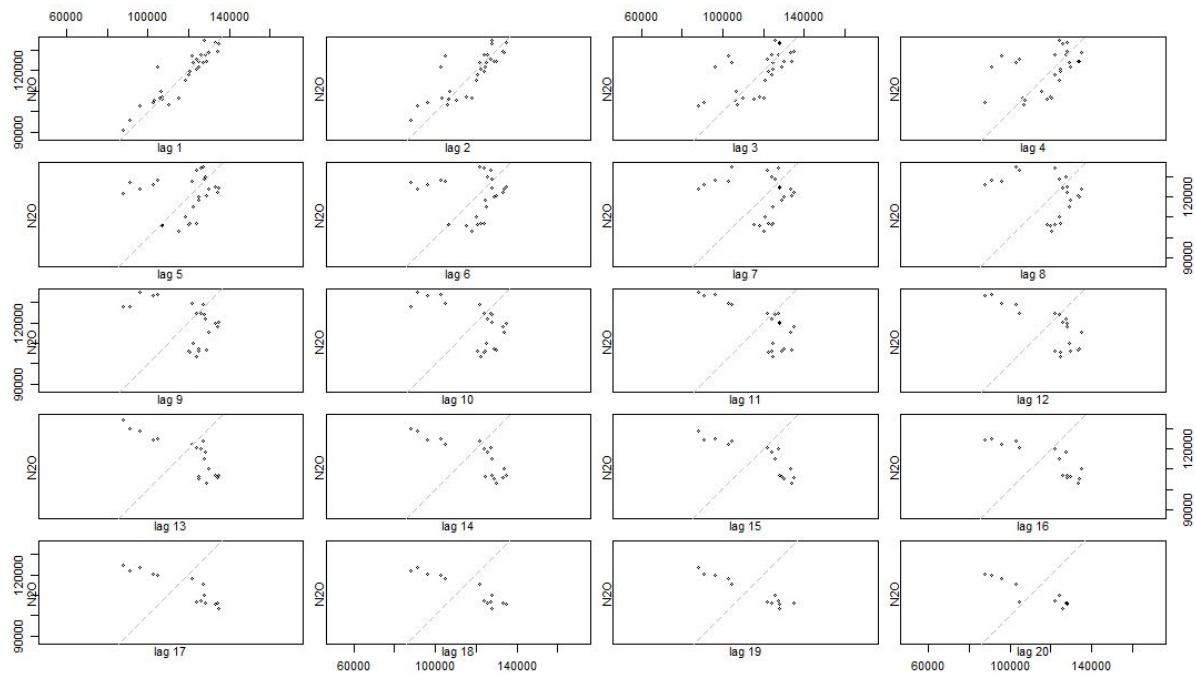
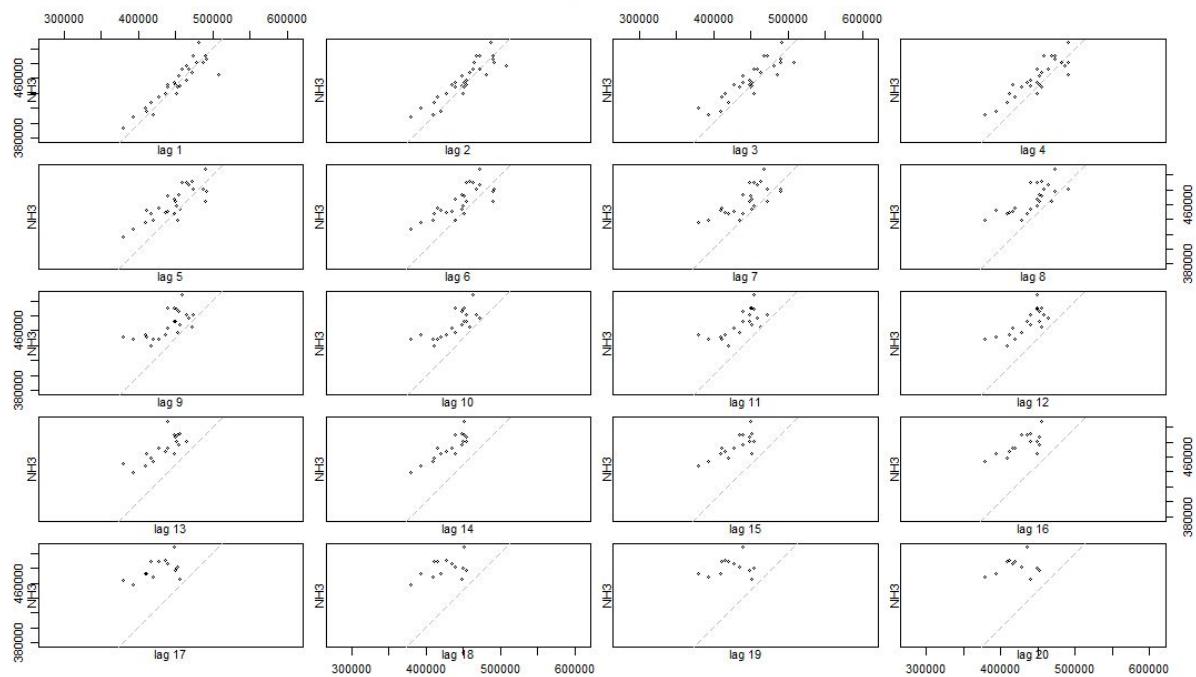
**Autodispersione COVNM****Autodispersione CH4**

## Autodispersione CO



## Autodispersione CO2



**Autodispersione N2O****Autodispersione NH3**

Osservazioni rilevanti sono quelli di N2O, CO2 e CH4 che presentano agglomerati di punti che tendono a diradarsi all'aumentare del lag tra i punti osservati.

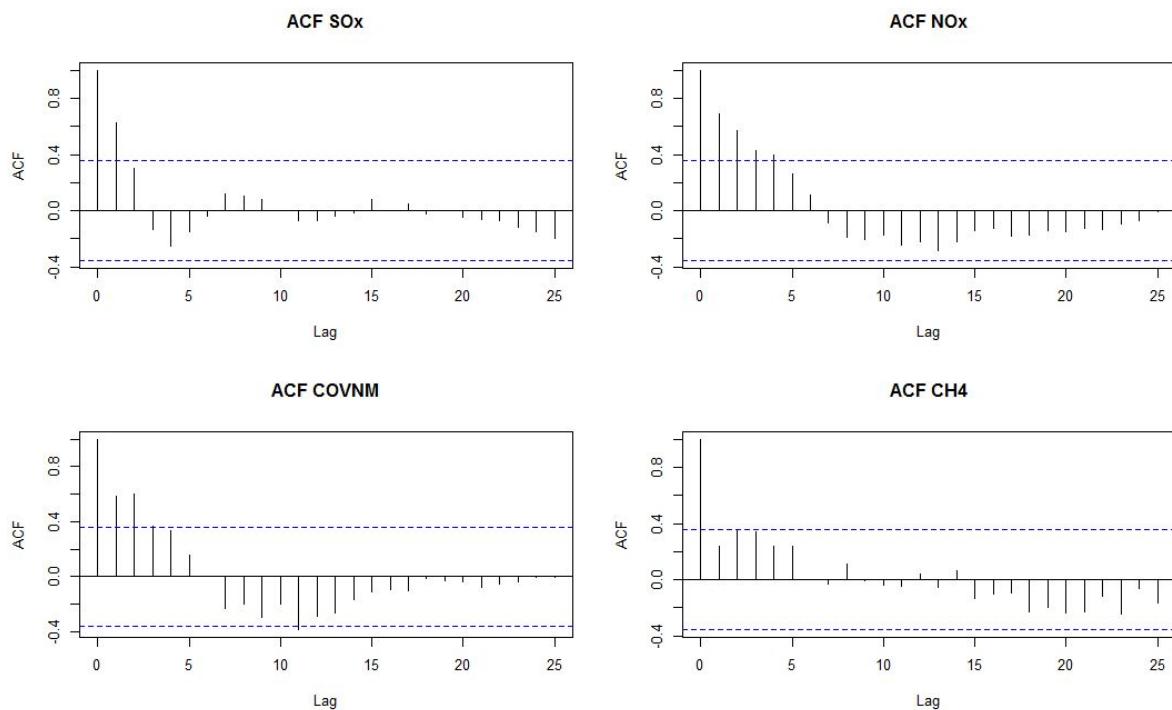
Per quanto riguarda il protossido di azoto ciò è possibile in quanto dal 2000 al 2005 ha avuto un andamento crescente legato alla crescita della produzione e nell'uso di fertilizzanti azotati. Successivamente con il decreto legislativo del 2006 la produzione e l'immissione in commercio dei fertilizzanti è stata limitata.

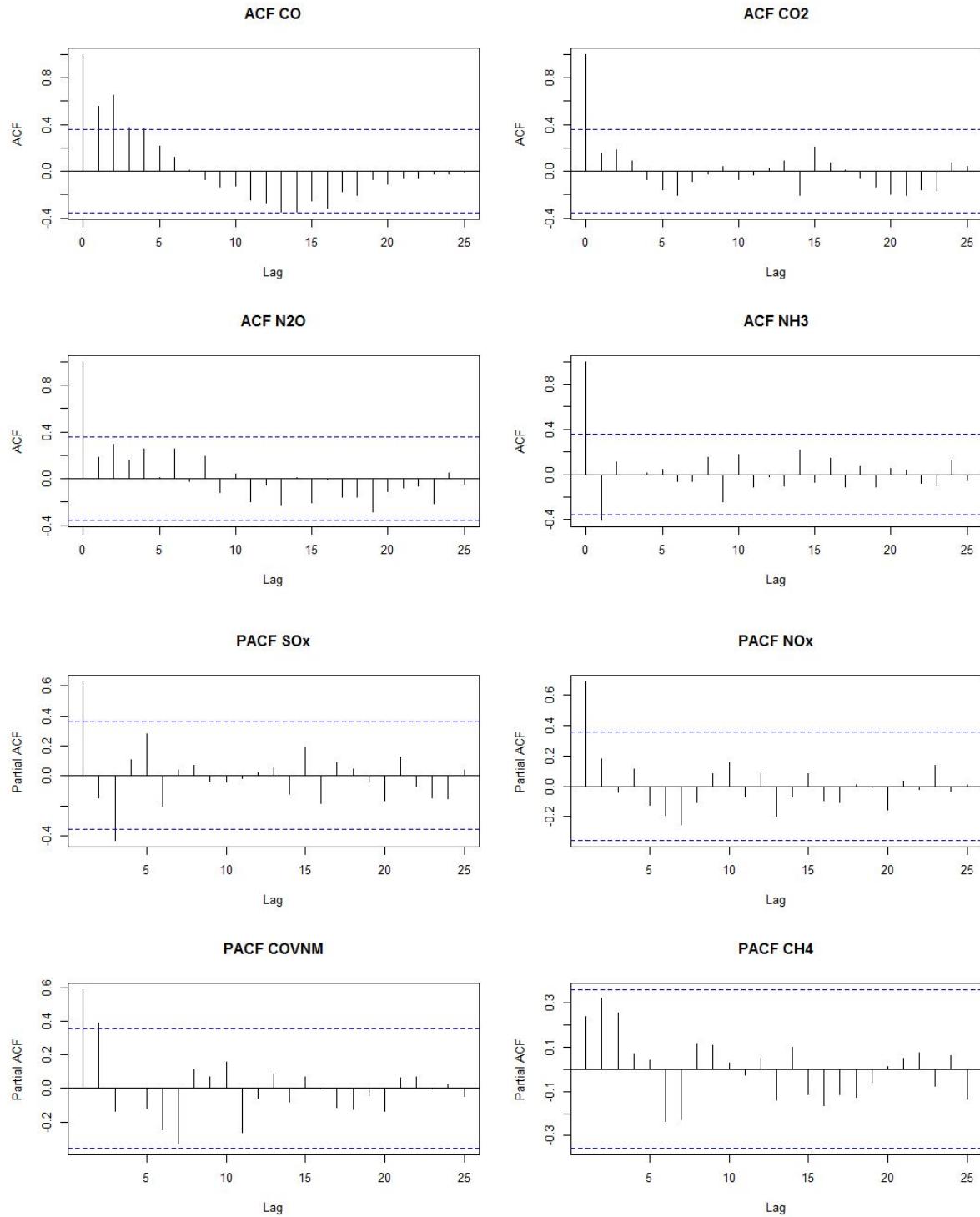
Mentre per il CO2 l'Italia, tra i paesi con economia avanzata, è sempre stato un basso emettitore, in quanto non è mai stata un grande utilizzatore del carbone per la produzione di energia elettrica preferendo ad esso gli idrocarburi. Anche il settore dei trasporti in passato ha avuto un ruolo importante nel mantenere basse le emissioni italiane, probabilmente per una minore diffusione degli autoveicoli e l'abitudine ad usare mezzi leggeri e quindi meno inquinanti. Tale virtuosità si è comunque annullata nel tempo portando un aumento il quale ha dunque costretto a prendere provvedimenti i quali tra il 2006 e 2008 hanno portato un rapido decremento.

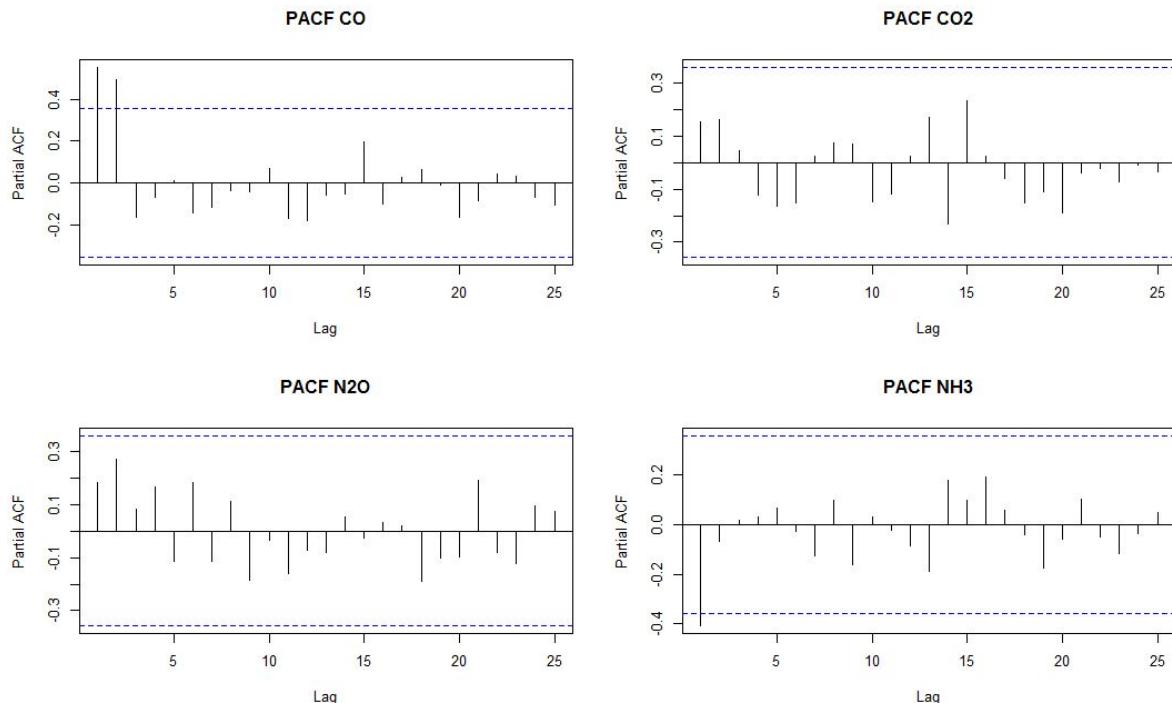
Il CH4 presenta in maniera più evidente una dilatazione tra i punti all'aumentare del lag, questa probabilmente è segnata dall'aumento delle emissioni fino agli inizi anni 2000 e poi successivamente un calo in seguito al decreto legislativo 2003.

L'ultima analisi fatta riguarda l'autocorrelazione globale e parziale a valle della detrendizzazione delle serie, tramite operatore differenza.

```
> diffSOx<-diff(sox, lag=1, differences = 1)
> diffNOx<-diff(NOX, lag=1, differences = 1)
> diffCOVNM<-diff(COVNM, lag=1, differences = 1)
> diffCH4<-diff(CH4, lag=1, differences = 1)
> diffCO<-diff(CO, lag=1, differences = 1)
> diffCO2<-diff(CO2, lag=1, differences = 1)
> diffN2O<-diff(N2O, lag=1, differences = 1)
> diffNH3<-diff(NH3, lag=1, differences = 1)
```







Dopo la detrendizzazione, i grafici di autocorrelazione suggeriscono che i residui delle serie CH4, SOX, CO, CO2, N2O, NH3 potrebbero appartenere ad un processo simile ad un WN in quanto vi è una scarsa autocorrelazione nel processo residuo. Esaminiamo a tal scopo i risultati del test di Ljung per le serie detrendizzate nei due modi visti (media mobile di ordine 2 e differenziazione).

```
> Box.test(detren_SOX, type = "Ljung", lag=25)
```

```
Box-Ljung test

data: detren_SOX
X-squared = 28.067, df = 25, p-value = 0.3048

> Box.test(diffSOX, type = "Ljung", lag=25)

Box-Ljung test

data: diffSOX
X-squared = 36.952, df = 25, p-value = 0.05834

> Box.test(detren_NOX, type = "Ljung", lag=25)

Box-Ljung test

data: detren_NOX
X-squared = 31.617, df = 25, p-value = 0.1694

> Box.test(diffNOX, type = "Ljung", lag=25)

Box-Ljung test

data: diffNOX
X-squared = 76.882, df = 25, p-value = 3.487e-07
```

```
> Box.test(detren_COVM, type = "Ljung", lag=25)
  Box-Ljung test
data: detren_COVM
X-squared = 28.183, df = 25, p-value = 0.2995

> Box.test(diffCOVNM, type = "Ljung", lag=25)
  Box-Ljung test
data: diffCOVNM
X-squared = 64.467, df = 25, p-value = 2.466e-05

> Box.test(detren_CH4, type = "Ljung", lag=25)
  Box-Ljung test
data: detren_CH4
X-squared = 18.957, df = 25, p-value = 0.7992

> Box.test(diffCH4, type = "Ljung", lag=25)
  Box-Ljung test
data: diffCH4
X-squared = 52.577, df = 25, p-value = 0.001013

> Box.test(detren_CO, type = "Ljung", lag=25)
  Box-Ljung test
data: detren_CO
X-squared = 12.581, df = 25, p-value = 0.9812

> Box.test(diffCO, type = "Ljung", lag=25)
  Box-Ljung test
data: diffCO
X-squared = 78.831, df = 25, p-value = 1.738e-07

> Box.test(detren_CO2, type = "Ljung", lag=25)
  Box-Ljung test
data: detren_CO2
X-squared = 37.307, df = 25, p-value = 0.05397

> Box.test(diffCO2, type = "Ljung", lag=25)
  Box-Ljung test
data: diffCO2
X-squared = 29.489, df = 25, p-value = 0.244
```

```
> Box.test(detren_N2O, type = "Ljung", lag=25)
  Box-Ljung test

data: detren_N2O
X-squared = 32.553, df = 25, p-value = 0.1426

> Box.test(diffN2O, type = "Ljung", lag=25)
  Box-Ljung test

data: diffN2O
X-squared = 40.059, df = 25, p-value = 0.02876

> Box.test(detren_NH3, type = "Ljung", lag=25)
  Box-Ljung test

data: detren_NH3
X-squared = 30.224, df = 25, p-value = 0.216

> Box.test(diffNH3, type = "Ljung", lag=25)
  Box-Ljung test

data: diffNH3
X-squared = 25.447, df = 25, p-value = 0.4376
```

Dai test risulta che le detrendizzazioni del CO e del CH4 fanno sì che i residui, ossia le serie detrendizzate, siano abbastanza vicine ad un WN, ciò non si può affermare per quanto riguarda la detrendizzazione fatta mediante differenziazione. Anche gli altri residui risultano essere lontani ad un WN questo è dovuto alle continue fluttuazioni temporali.