ANALISI INDICE NASDAQ MAGGIO 2012 - APRILE 2017

Table of Contents

1) Presentazione titolo azionario 1
2) Analisi preliminare grafici 1
3) Test DF per la verifica di stazionarietà 4
4) Analisi della distribuzione dei rendimenti 6
5) Analisi del correlogramma empirico 8
6) Stima del modello ARIMA 10

Michele Sordo: il caso della time serie dell'indice ^NDX100 del NYSE.

1) Presentazione titolo azionario

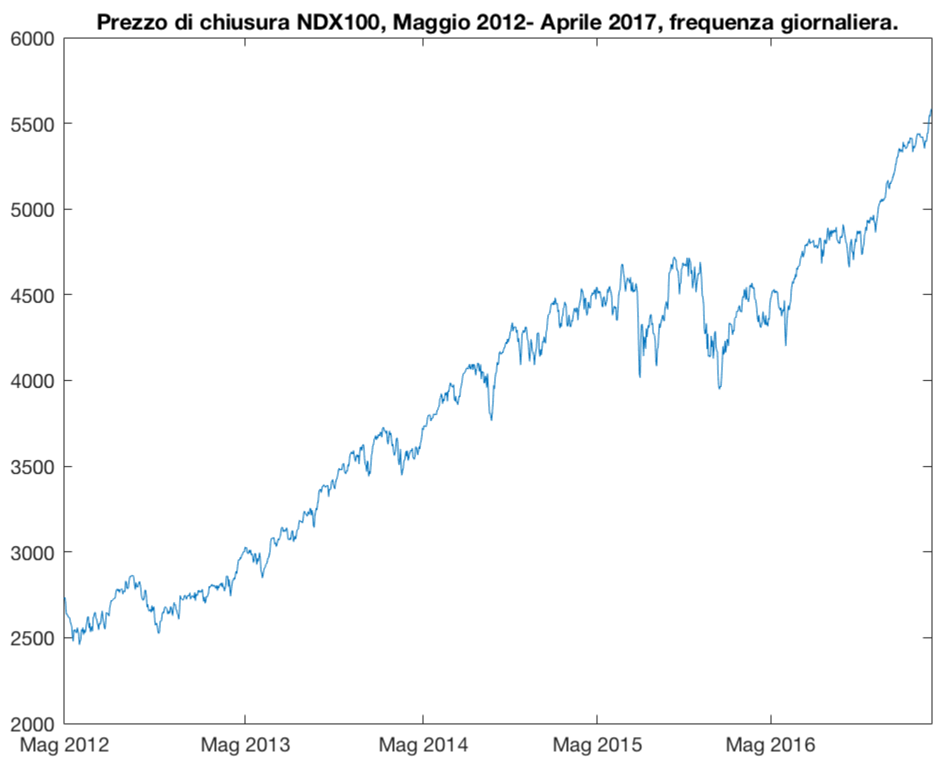
Il NASDAQ-100 è un indice di borsa delle maggiori 100 imprese non-finanziarie quotate nel mercato borsistico NASDAQ. È un indice ponderato; il peso delle diverse società che lo compongono è basato sulla loro capitalizzazione di mercato, con alcune regole per tener conto delle influenze delle componenti maggiori. Non comprende società finanziarie, e include alcune società estere. Questi due fattori lo differenziano dall'indice S&P 500. (Fonte: https://it.wikipedia.org/wiki/NASDAQ-100) La serie analizzata contiene dati a frequenza giornaliera degli ultimi 5 anni: dal 1 Maggio 2012 al 28 Marzo 2017 è presente un totale di 1257 osservazioni. L'indice è quotato 5 giorni a settimana, dal lunedì al venerdì.

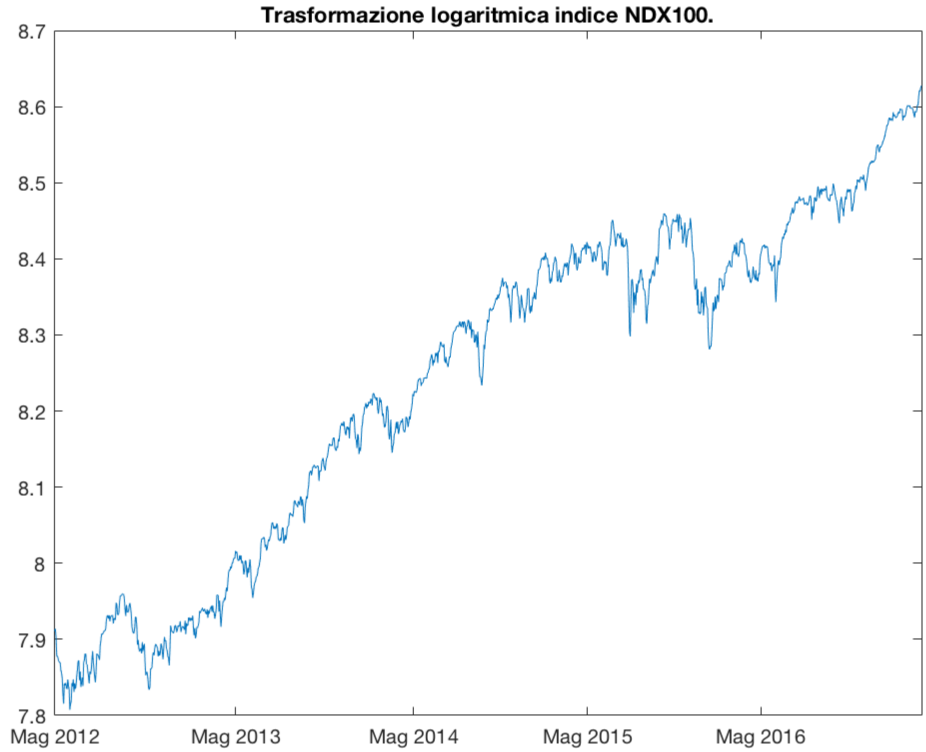
2) Analisi preliminare grafici

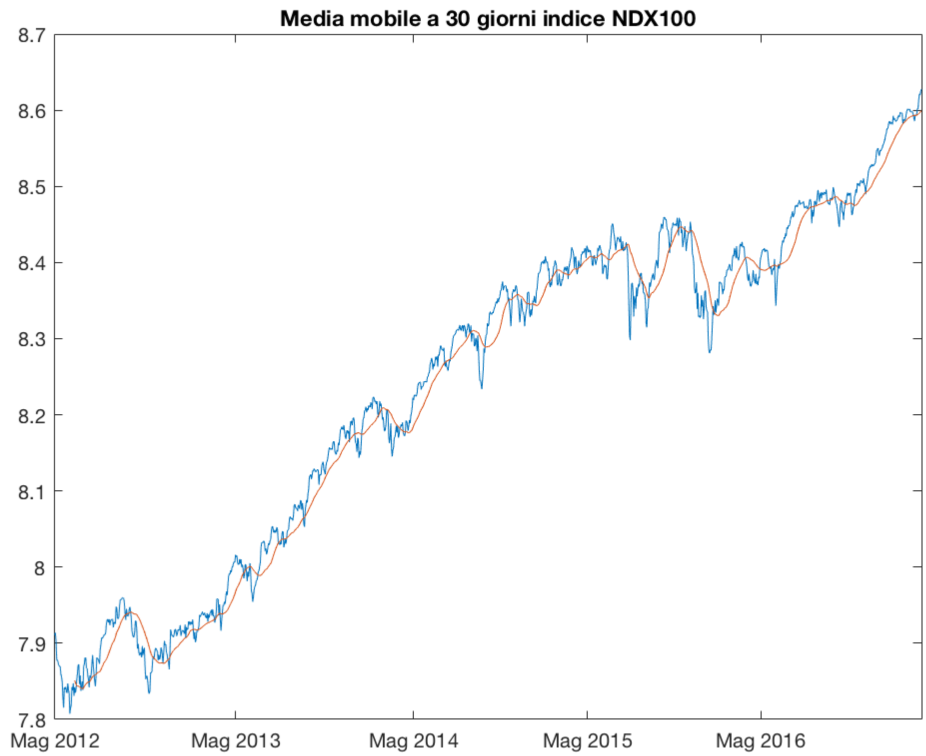
Analisi preliminare della serie storica.

```
load('NDX.mat');             % caricamento dei dati nel workspace
y = NDX;                     % definizione della serie
T = length(y);               % definizione di T numero di osservazioni
t = (1:T);                   % definizione del vettore temporale


figure;                      % apertura nuova figura vuota
plot(y);                     % plot grafico della serie
h1 = gca;                    % definizione assi cartesiani
h1.XLim = [0,T];             % definizione asse delle ascisse
h1.XTick = [1 263 521 773 1024];   % definizione etichette asse
 ascisse
h1.XTickLabel = {'Mag 2012','Mag 2013','Mag 2014','Mag 2015', 'Mag
 2016'};  % definizione titolo grafico
title('Prezzo di chiusura NDX100, Maggio 2012- Aprile 2017, frequenza
 giornaliera.');

ly = log(y);                 % definizione serie trasformata
 logaritmica
```

```
figure;
plot(ly);
h1 = gca;
h1.XLim = [0,T];
h1.XTick = [1 263 521 773 1024];
h1.XTickLabel = {'Mag 2012','Mag 2013','Mag 2014','Mag 2015', 'Mag
 2016'};
title('Trasformazione logaritmica indice NDX100.');

ma = tsmovavg(ly,'s',30,1);     % calcolo media mobile a 30 giorni
figure;
plot(t,ly,t,ma);
h1 = gca;
h1.XLim = [0,T];
h1.XTick = [1 263 521 773 1024];
h1.XTickLabel = {'Mag 2012','Mag 2013','Mag 2014','Mag 2015', 'Mag
 2016'};
title('Media mobile a 30 giorni indice NDX100');
```




Trasformazione logaritmica indice NDX100.
8.7
8.6
8.5
8.4
8.3
8.2
8.1
8
7.9
7.8
Mag 2012
Mag 2013
Mag 2014
Mag 2015
Mag 2016


Media mobile a 30 giorni indice NDX100
8.7
8.6
8.5
8.4
8.3
8.2
8.1
8
7.9
7.8
Mag 2012
Mag 2013
Mag 2014
Mag 2015
Mag 2016

I primi due grafici illustrano rispettivamente l'andamento della serie storica dell'indice NDX100 e della sua rispettiva trasformata logaritmica. Il terzo grafico è utilizzato per indicare l'andamento della media mobile a trenta giorni della trasformata logaritmica, strumento utilizato per l'analisi degli investimenti borsistici. La media mobile è calcolata sommando i rispettivi ritardi delle osservazioni, nel nostro caso 30, e dividendo la sommatoria per il numero delle osservazioni. In questo modo, di giorno in giorno i valori superiori a 30 vengono sostituiti dalle nuove misurazioni, indicando un andamento della serie.

L'indicatore di media mobile viene utilizzato nell'ambito dell'analisi tecnica per evidenziare le tendenze sottostanti l'andamento dei prezzi eliminando l'effetto di fluttuazioni minori (transitorie). Le medie mobili possono essere calcolate con diverse metodologie. Le principali metodologie di calcolo sono le seguenti: medie mobili aritmetiche, medie mobili esponenziali e medie mobili ponderate. Eventuali cambiamenti di direzi%one della media mobile, oppure un incrocio tra la media mobile e la serie dei prezzi possono rappresentare segnali di acquisto o vendita, mentre la media mobile in sè rappresenta spesso un supporto o una resistenza. Fonte (http://www.borsaitaliana.it/bitApp/glossary.bit?target=GlossaryDetail&word=Media%20Mobile). Questa prima analisi grafica permette di rilevare la mancanza di stazionarietà in senso debole della serie, in quanto la media presenta picchi anche distanti tra loro e la varianza si presenta contenuta in determinati periodi e più alta in altri. Al fine di verificare questa ipotesi utilizziamo il test di Dickey-Fuller.

3) Test DF per la verifica di stazionarietà

Il test di Dickey-Fuller verifica l'ipotesi nulla di presenza di una radice unitaria all'interno del processo generatore della serie storica. Se l'ipotesi nulla non viene rifiutata è necessario trasforamre la serie, al fine di eliminare l'effetto della radice unitaria. Se l'ipotesi nulla viene rifiutata la serie può essere considerata stazionaria, a meno di particolari fattori. Matlab include di default 20 ritardi della variabile dipendente nella regressione ausiliaria utilizzata dalla funzione.

```
[h,pValue,stat] = adftest(ly);      % calcolo della statistica DF
table(h,stat,pValue)                % output del test
if h == 1
    fprintf('Rifiuto l''ipotesi nulla');
else
    fprintf('Non rifiuto l''ipotesi nulla');
end                                 % esito del test
```

```
ans =

      h        stat       pValue
    _____    ______     _______

    false    2.1213     0.99194

Non rifiuto l'ipotesi nulla
```

L'output della funzione mostra rispettivamente: del test mostra rispettivamente

- H: l'esito dell'ipotesi alternativa (true: rifiuto l'ipotesi nulla, false: non rifiuto l'ipotesi nulla);
- STAT: valore statistica test associata;
- P-VALUE: p-value associato.

Nel nostro caso l'esito è "h = False", quindi non posso rifiutare l'ipotesi nulla di presenza di radice unitaria nella serie storica. Di conseguenza calcolo la serie delle differenze prime e ripeto il test per verificarne la stazionarietà in senso debole.

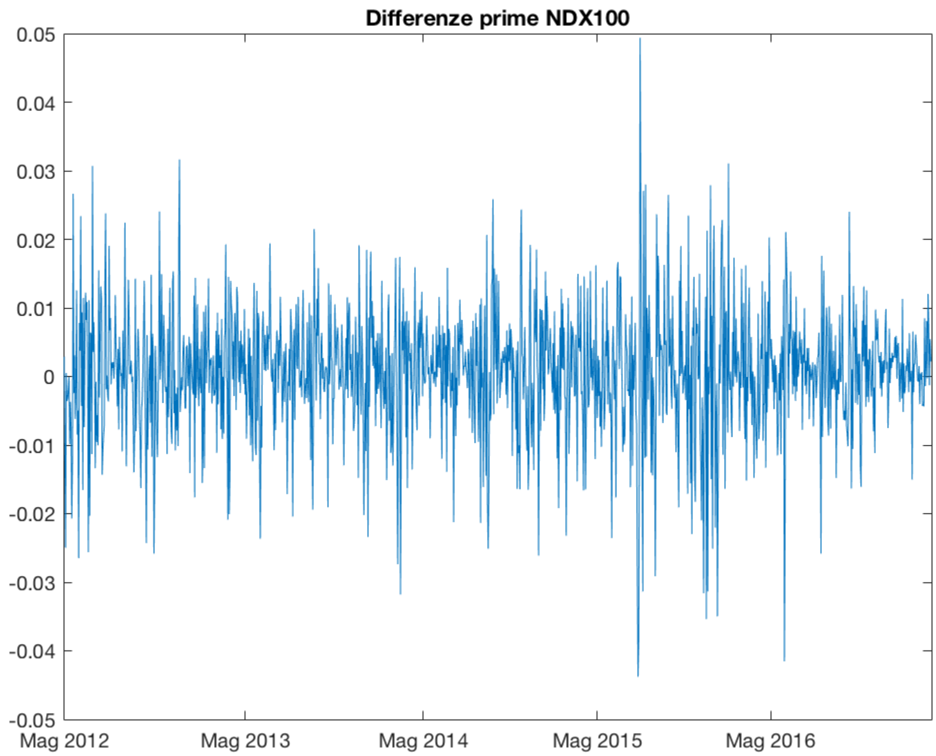
```
dy = diff(ly);                              % calcolo delle differenze prime
figure;
plot(dy);
h2 = gca;
h2.XLim = [0,T];
h2.XTick = [1 263 521 773 1024];
h2.XTickLabel = {'Mag 2012','Mag 2013','Mag 2014','Mag 2015', 'Mag
 2016'};
title('Differenze prime NDX100');

[h,pValue,stat] = adftest(dy);
table(h,stat,pValue)
if h == 1
    fprintf('Rifiuto l''ipotesi nulla');
else
    fprintf('Non rifiuto l''ipotesi nulla');
end
```

Warning: Test statistic #1 below tabulated critical values:
minimum p-value = 0.001 reported.

ans =

h	*stat*	*pValue*
true	*-34.491*	*0.001*

Rifiuto l'ipotesi nulla



L'esito del test Dickey-Fuller, in questo caso, è positivo quindi rifiuto l'ipotesi di presenza di una seconda radice unitaria e traggo la conclusione che la serie storica dell'indice NDX100 si possa considerare stazionaria in senso debole nelle differenze prime. Il grafico associato alla serie delle differenze prime conferma che la serie è almeno stazionaria in media e mostra che la varianza cambia nei diversi periodi, risultando più marcata in alcuni casi e più contenuta in altri. Al fine dell'analisi della variabilità verrà affrontata nella sezione 8. Consideriamo quindi la serie come "integrata di ordine 1", ossia I(1).

4) Analisi della distribuzione dei rendimenti

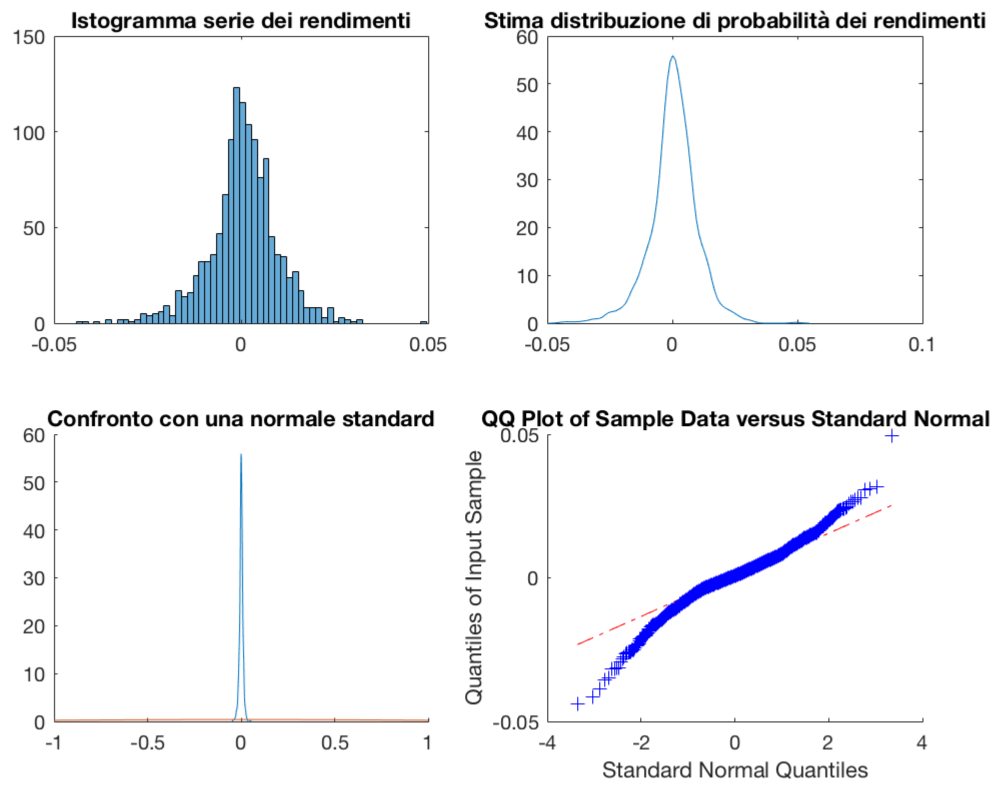
La serie delle differenze prime corrisponde alla serie dei rendimenti dell'indice NDX100. Al fine di scegliere e stimare il modello migliore si utilizza l'approccio di Box-Jenkins, definendo il processo stocastico di generazione della serie storica.

```
Media = mean(dy);
Varianza = var(dy);
Curtosi = kurtosis(dy);
Asimmetria = skewness(dy);
table(Media, Varianza, Curtosi, Asimmetria)
figure
subplot(2,2,1)
histogram(dy,60)
title('Istogramma serie dei rendimenti')

[f,dyi] = ksdensity(dy);
subplot(2,2,2)
plot(dyi,f)
```

```
title('Stima distribuzione di probabilità dei rendimenti')

subplot(2,2,3)
hold on
plot(dyi,f)
x = (-1:.1:1);
norm = normpdf(x,0,1);
plot(x,norm)
hold off
title('Confronto con una normale standard')

subplot(2,2,4)
qqplot(dy)
```

ans =

Media	*Varianza*	*Curtosi*	*Asimmetria*
0.00057058	*8.9686e-05*	*5.2899*	*-0.32627*



I momenti calcolati mostrano:

- una media prossima allo zero, pari a "0.00057058".

- una varianza prossima allo zero, pari a "8.9686e-05".

La distribuzione è leptocurtica e asimmetrica verso sinistra. Il confronto tra la stima della distrubuzione di probabilità dei rendimenti ed il confronto con la normale standard mostra che, sebbene le osservazioni siano concentrate intorno alla media, la distribuzione presenta pesanti code da entrambe le parti, con una particolare accentuazione verso sinistra. L'analisi è confermata dal qq-plot, dove si nota la presenza di asimmetria e la conferma dell'ipotesi delle code più pesanti.

5) Analisi del correlogramma empirico

Al fine di identificare il processo generatore della serie tramite l'approccio di Box-Jenkins si calcolano e si rappresentano attraverso i rispettivi correlogrammi empirici:

- ACF: funzione di autocorrelazione

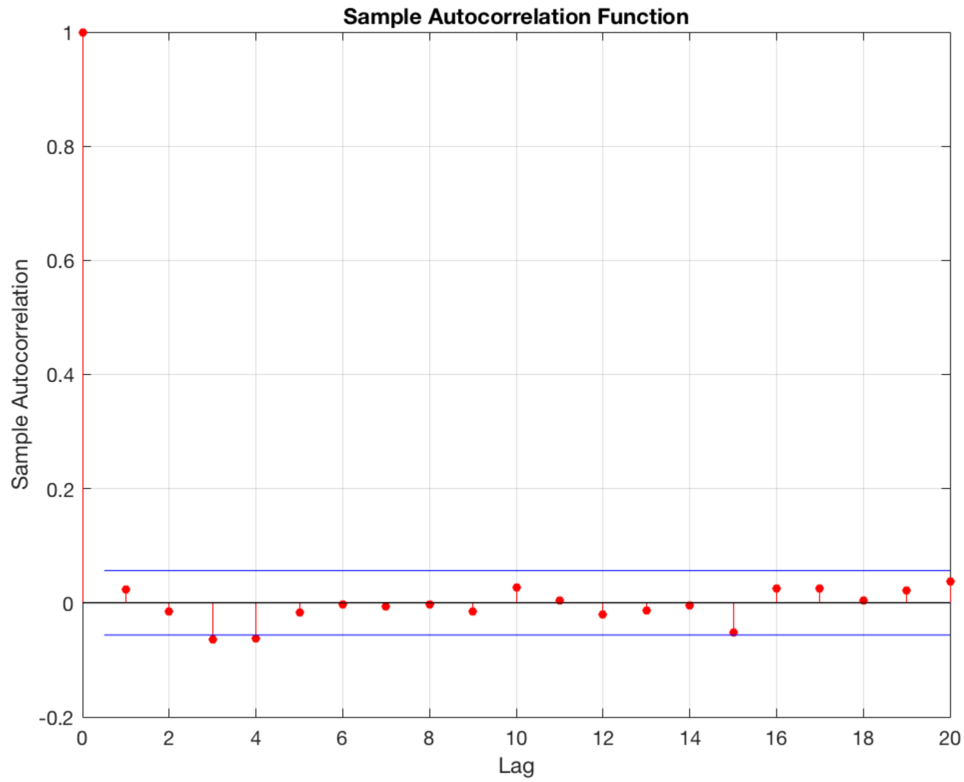
- PACF: funzione di autocorrelazione parziale.

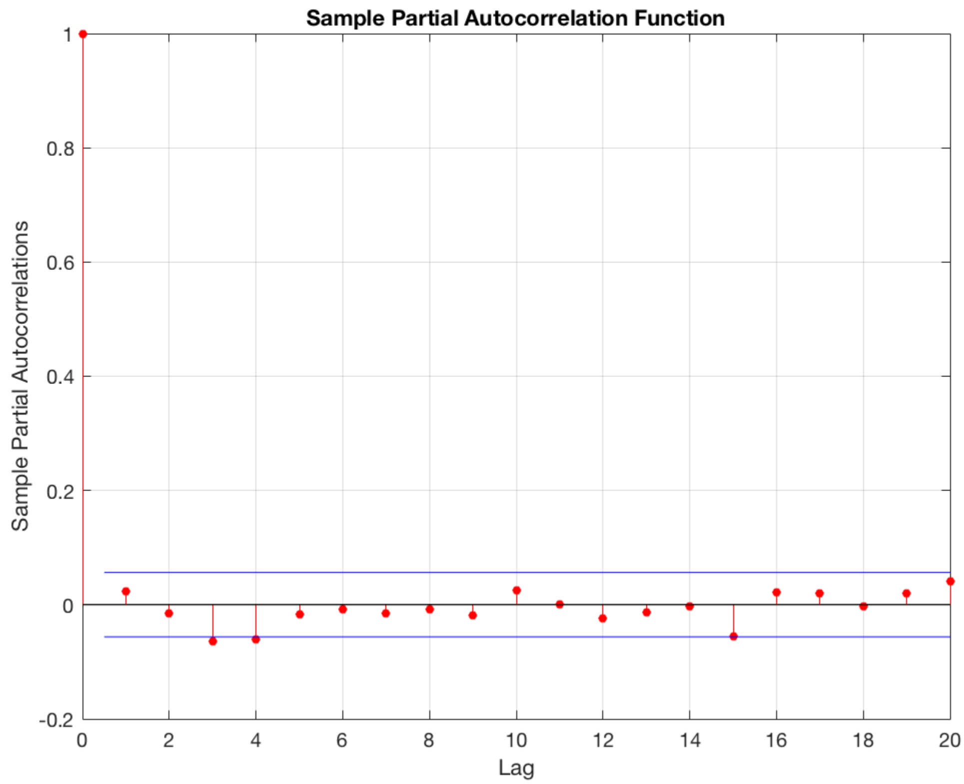
I correlogrammi empirici presentanto una prima analisi grafica della relazione tra la serie e le osservazioni.

```
K = (0:20)';                     % definizione vettore ritardi
ACF = autocorr(dy);              % calcolo ACF
PACF = parcorr(dy);              % calcolo PACF
table(K,ACF,PACF)
figure
autocorr(dy)                     % plot correlogramma ACF
figure
parcorr(dy)                      % plot correlogramma PACF
```

```
ans =

    K        ACF            PACF
    __    __________    ___________

     0             1              1
     1      0.022853       0.022853
     2     -0.015352       -0.01589
     3     -0.064379      -0.063775
     4      -0.06288      -0.060681
     5     -0.016464      -0.016268
     6    -0.0028925     -0.0085681
     7    -0.0064258      -0.015068
     8    -0.0023477      -0.008392
     9       -0.0155      -0.018985
    10      0.026972        0.02482
    11     0.0038611    -0.00017703
    12        -0.021      -0.024464
    13     -0.013686      -0.013015
    14    -0.0046123     -0.0022273
    15      -0.05162      -0.054756
    16      0.024571       0.021842
    17      0.025596       0.020668
    18     0.0048912     -0.0036883
```

19	*0.021881*	*0.019745*
20	*0.037173*	*0.040337*





I correlogrammi empirici non permettono di derivare direttamente il modello stocastico migliore per l'analisi della serie storica ma permettono di notare i potenziali effetti dei ritardi, osservando per quali K essi superino l'intervallo di confidenza. Nel nostro caso i ritardi che mostrano potenziali effetti sono il terzo e il quarto. Tramite questa informazioni si procede a stimare il modello ARIMA migliore e più parsimonioso.

6) Stima del modello ARIMA

```
Mdl = arima(1,1,0);           % definizione del modello ARIMA(1,1,0)
EstMdl10 = estimate(Mdl,ly);  % stima del modello definito sui dati
 della serie storica

Mdl = arima(0,1,1);             % ripetizione del procedimento
EstMdl01 = estimate(Mdl,ly);

Mdl = arima(1,1,1);
EstMdl11 = estimate(Mdl,ly);
```

```
    ARIMA(1,1,0) Model:
    --------------------
    Conditional Probability Distribution: Gaussian

                                  Standard          t
     Parameter       Value          Error        Statistic
    -----------   -----------   ------------   -----------
     Constant     0.000555734   0.000271965       2.04341
```

```
       AR{1}       0.0228532      0.0206712         1.10556
    Variance     8.97064e-05    2.48051e-06         36.1645

    ARIMA(0,1,1) Model:
    --------------------
    Conditional Probability Distribution: Gaussian

                                   Standard          t     
     Parameter       Value          Error        Statistic 
    -----------   -----------   ------------   -----------
     Constant     0.000567465     0.0002787         2.03611
        MA{1}       0.0236948     0.0206558         1.14713
     Variance     8.94659e-05   2.46873e-06         36.2396

    ARIMA(1,1,1) Model:
    --------------------
    Conditional Probability Distribution: Gaussian

                                   Standard          t     
     Parameter       Value          Error        Statistic 
    -----------   -----------   ------------   -----------
     Constant     0.000608959   0.000559077         1.08922
        AR{1}      -0.0725011      0.970821      -0.0746802
        MA{1}       0.0962038      0.970265       0.0991521
     Variance     8.94876e-05   2.48987e-06         35.9406
```

```
Mdl = arima(2,1,1);
EstMdl = estimate(Mdl,ly);
res = infer(EstMdl,ly);

Mdl = arima('ArLags',[3 4],'D',1,'MaLags',[]);
EstMdl340 = estimate(Mdl,ly);


Mdl = arima('ArLags',[],'D',1,'MaLags',[3 4]);
EstMdl034 = estimate(Mdl,ly);


Mdl = arima('ArLags',[3 4],'D',1,'MaLags',[3 4]);
EstMdl3434 = estimate(Mdl,ly);
```

```
    ARIMA(2,1,1) Model:
    --------------------
    Conditional Probability Distribution: Gaussian

                                   Standard          t     
     Parameter       Value          Error        Statistic 
    -----------   -----------   ------------   -----------
     Constant     0.000158659   0.000115054         1.37899
        AR{1}        0.793469      0.125466         6.32418
        AR{2}      -0.0655226     0.0218715         -2.9958
        MA{1}       -0.775071      0.125727         -6.1647
     Variance     8.89974e-05   2.53921e-06         35.0493
```

```
    ARIMA(4,1,0) Model:
    --------------------
    Conditional Probability Distribution: Gaussian

                                    Standard          t
     Parameter       Value           Error        Statistic
    -----------   -----------   ------------   -----------
     Constant      0.0006686    0.000274536        2.43538
        AR{3}     -0.0629839      0.0222096       -2.83589
        AR{4}     -0.0614709      0.0223005       -2.75648
     Variance    8.86253e-05    2.50313e-06        35.4058

    ARIMA(0,1,4) Model:
    --------------------
    Conditional Probability Distribution: Gaussian

                                    Standard          t
     Parameter       Value           Error        Statistic
    -----------   -----------   ------------   -----------
     Constant    0.000569214    0.000245826        2.31552
        MA{3}     -0.0644678      0.0226558       -2.84553
        MA{4}     -0.0635488      0.0231999       -2.73918
     Variance    8.88093e-05    2.50972e-06        35.3862

    ARIMA(4,1,4) Model:
    --------------------
    Conditional Probability Distribution: Gaussian

                                    Standard          t
     Parameter       Value           Error        Statistic
    -----------   -----------   ------------   -----------
     Constant    0.000146416    4.48184e-05        3.26687
        AR{3}       0.190689       0.119548        1.59508
        AR{4}       0.579912       0.109864        5.27847
        MA{3}        -0.2421       0.115203       -2.10151
        MA{4}      -0.634331       0.105668       -6.00307
     Variance    8.81666e-05    2.45867e-06        35.8594

%{


%% 7) Analisi dei residui
%Valutiamo la qualità del modello specificato analizzando la
 distribuzione dei residui della regressione.

sy = (dy-res(2:end));
figure
h1 = gca;
h1.XLim = [0,T];
h1.XTick = [1 263 521 773 1024];
hold on
plot(dy)
plot(sy,'r')
```

```
legend('Valori osservati','Valori stimati')
title('Differenza tra valori osservati e stimati')
hold off

figure
plot(res./sqrt(EstMdl.Variance))        % plot dei residui
 standardizzati
h1 = gca;
h1.XLim = [0,T];
h1.XTick = [1 263 521 773 1024];
title('Residui standardizzati')

%%


figure
subplot(2,2,1)
histogram(res,60)                       % istogramma dei residui
title('Istogramma residui')
subplot(2,2,2)
qqplot(res)                             % qqplot dei residui
subplot(2,2,3)
autocorr(res)                           % correlogramma
subplot(2,2,4)
parcorr(res)

%%


[h,p,Qstat,crit] = lbqtest(res,'lags',[1 2 3 4 5 6 7 8 9 10 11 12 13
 14 15 16 17 18 19 20]);
names = {'K';'H';'pvalue';'Qstat';'Crit'};
table(K(2:end),h',p',Qstat',crit','VariableNames',names)        %
 creazione tabella ouput



%% 8) Test per la presenza di effetti ARCH

figure;
plot((res-mean(res)).^2);
h1 = gca;
h1.XLim = [0,T];
h1.XTick = [1 263 521 773 1024];
figure;
subplot(2,1,1)
autocorr((res).^2);
subplot(2,1,2)
parcorr((res).^2);
[h,p,Qstat,crit] = lbqtest(res.^2,'lags',1);
table(h,Qstat,crit,p,'rownames',{'lbqtest'})
if h == 1
    fprintf('Rifiuto l''ipotesi nulla');
else
```

```
    fprintf('Non rifiuto l''ipotesi nulla');
end
[h,p] = archtest(res,'lags',1);     %Test di Engle
table(h,p,'rownames',{'Archtest'})
if h == 1
    fprintf('Rifiuto l''ipotesi nulla');
else
    fprintf('Non rifiuto l''ipotesi nulla');
end




%% 9) Selezione e stima modello GARCH

Mdl = egarch(0,1);
VarMdl = estimate(Mdl,res);

%%
%
%
%
%
%

cv = infer(VarMdl,res);     % generazione della varianza condizionata
stres = (res./sqrt(cv));    % standardizzazione dei residui
[h,p,Qstat,crit] = lbqtest(cv.^2,'lags',2);
table(h,Qstat,crit,p,'rownames',{'lbqtest'})
if h == 1
    fprintf('Rifiuto l''ipotesi nulla');
else
    fprintf('Non rifiuto l''ipotesi nulla');
end
[h,p] = archtest(cv,'lags',2);         %Test di Engle
table(h,p,'rownames',{'Archtest'})
if h == 1
    fprintf('Rifiuto l''ipotesi nulla');
else
    fprintf('Non rifiuto l''ipotesi nulla');
end

%%
%
%
%
%
%

figure;
subplot(2,1,1)

title('ACF residui standardizzati al quadrato');
subplot(2,1,2)
```

```
parcorr((stres).^2);
title('PACF residui standardizzati al quadrato');
figure
subplot(2,2,1)
plot(stres)
h1 = gca;
h1.XLim = [0,T];
h1.XTick = [1 263 521 773 1024];
title('Residui standardizzati');
subplot(2,2,2)
qqplot(stres)
subplot(2,2,3)
autocorr(stres)
subplot(2,2,4)
parcorr(stres)
figure
h1 = gca;
h1.XLim = [0,T];
h1.XTick = [1 263 521 773 1024];
hold on
plot(res./sqrt(EstMdl.Variance))
plot(stres,'r');
hold off
title('Residui standardizzati con varianza EGARCH su residui
 standardizzati con varianza costante');
figure
h1 = gca;
h1.XLim = [0,T];
h1.XTick = [1 263 521 773 1024];
scatter(stres,res);
ylabel('Residui originali');
xlabel('Residui standardizzati');
title('Scatter plot residui standardizzati e residui originali')
autocorr((stres).^2);


%% 10) Previsione

[yf, vf] = forecast(EstMdl,30,'Y0',ly(1:1247));         % valore
 atteso condizionato
figure
h1 = gca;
h1.XLim = [0,T];
h1.XTick = [1 263 521 773 1024];
h1.XTickLabel = {'Giu 2011','Giu 2012','Giu 2013','Giu 2014','Giu
 2015'};
hold on
h2 = plot(ly,'Color',[.7,.7,.7]);
h3 = plot(1248:1277,yf,'b','LineWidth',2);
h4 = plot(1248:1277,yf + 1.96*sqrt(vf),'r:','LineWidth',2);
plot(1248:1277,yf - 1.96*sqrt(vf),'r:','LineWidth',2);
legend([h2 h3 h4],'Valori osservati','Previsione','Intervallo di
 confidenza al 95%','Location','NorthWest');
```

```
title('Simulazione previsione per 30 periodi e intervalli di
 confidenza al 95%')
hold off

%%
%
%
%
%
%

cvf = forecast(VarMdl,30,'Y0',res(1:1247));         % varianza
 condizionata
figure
h1 = gca;
h1.XLim = [0,T];
h1.XTick = [1 263 521 773 1024];
h1.XTickLabel = {'Giu 2011','Giu 2012','Giu 2013','Giu 2014','Giu
 2015'};
hold on
plot(1:1247,cv(1:1247),'b','Color',[.7,.7,.7]);
plot(1248:1277,cvf,'r');
legend('Varianza condizionata','Previsione','Location','NorthEast');
title('Simulazione previsione varianza condizionata EGARCH(0,1) per 30
 periodi');

%%
%
%
%
%
%

Mdl = arima('ArLags',[1 5],'MaLags',[1 5 6]);  % simulazione
 rendimenti con varianza EGARCH
EstMdl2 = estimate(Mdl,dy);
yf2 = forecast(EstMdl2,100,'Y0',dy(1:1177));
cvf2 = forecast(VarMdl,100,'Y0',res(1:1177));
figure
h1 = gca;
h1.XLim = [0,T];
h1.XTick = [1 263 521 773 1024];
h1.XTickLabel = {'Giu 2011','Giu 2012','Giu 2013','Giu 2014','Giu
 2015'};
hold on
h2 = plot(dy,'Color',[.7,.7,.7]);
h3 = plot(1178:1277,yf2,'b','LineWidth',2);
h4 = plot(1178:1277,yf2 + 1.96*sqrt(cvf2),'r:','LineWidth',2);
plot(1178:1277,yf2 - 1.96*sqrt(cvf2),'r:','LineWidth',2);
legend([h2 h3 h4],'Valori osservati','Previsione','Intervalli di
 confidenza al 95%','Location','NorthWest');
title('Simulazione previsione RENDIMENTI per 100 periodi e intervalli
 di confidenza al 95%')
hold off
```

```
%% 11) Previsioni

close all

%}
```

Published with MATLAB® R2016b