
Table of Contents

Lab Sheet 7: Forecasting ARMA and Reg-ARMA	1
Q1 (forecasting)	1
Q1(a) Exploratory Data Analysis	3
Q1(b) Determine forecast models and provide initial forecasts	4
Q1(c) Moving origin forecasts	12

Lab Sheet 7: Forecasting ARMA and Reg-AR-MA

Import AORD, TLS and BHP daily data from files "AllORD00-17.csv", "BHP00-17.csv" and "TLS00-17.csv" respectively I import these datasets as 'matrices' called AOdata, BHPdata and TLSdata respectively (all 7 columns). I also separately import just the date columns (first columns) as a 'column vectors' called AOdates, BHPdates and TLSdates respectively in Datenum format. save lab7.mat; Probably worth saving data at this stage so you need not repeat somewhat laborious import steps later!

```
load lab7.mat;
```

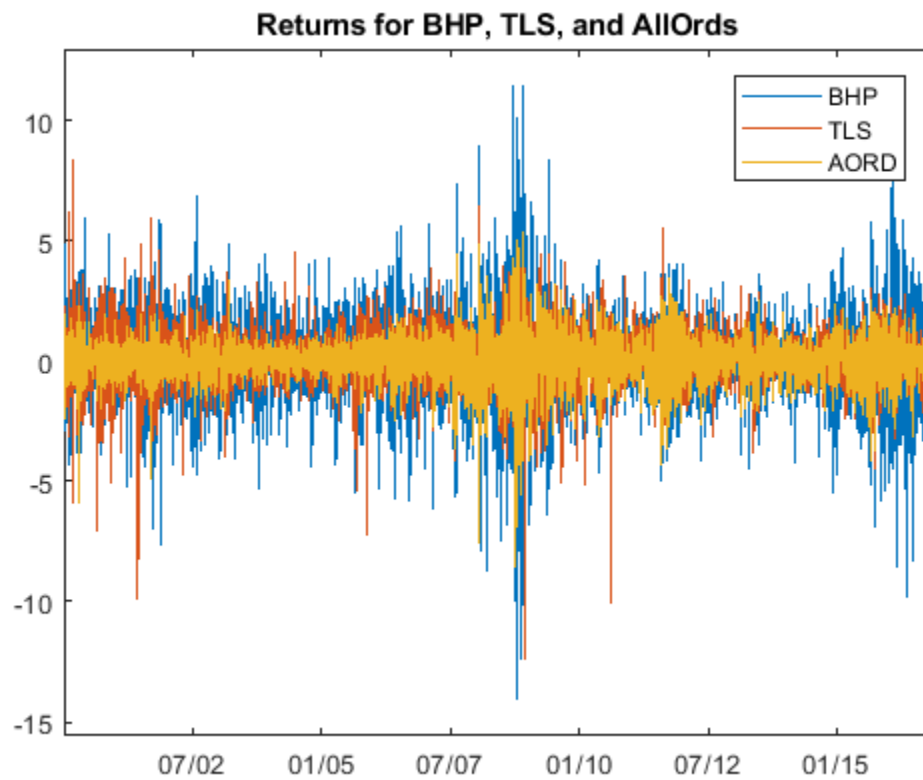
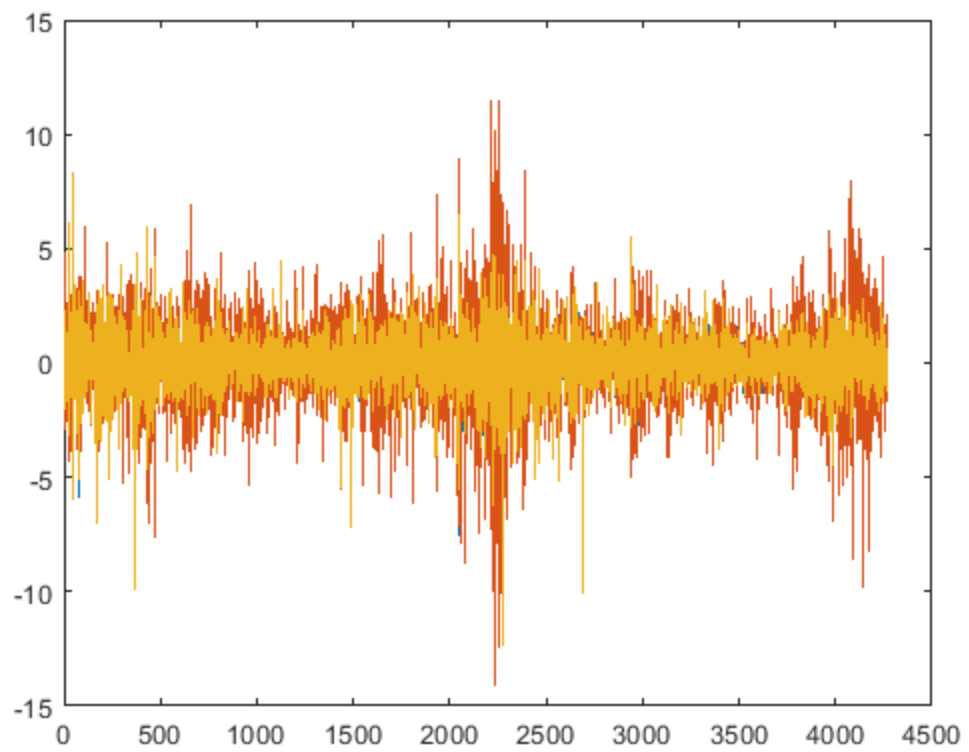
Q1 (forecasting)

```
% fadat function matches dates across data sets and converts to log
% returns
[yrt,ydat]=fadat(AOdata,BHPdata,TLSdata,AOdates,BHPdates,TLSdates);
% Note: fadat assumes the price data is in column 7 of the numeric
% matrices
% yrt has AO, BHP, and TLS returns in cols 1, 2, and 3 respectively

clear AOdata BHPdata TLSdata AOdates BHPdates TLSdates;

figure;plot(yrt); % basic plot with no formatting

figure;plot(datenum(ydat),[yrt(:,2) yrt(:,3) yrt(:,1)]); % include
% dates and change order so AOrd is plotted last and is then on top
legend('BHP','TLS','AORD','location','northeast'); % add legend
datetick('x','mm/yy'); % sets format for axis labels and preserves
% ticks and limits and keeps them in this format
axis([min(datenum(ydat)) max(datenum(ydat)) min((min(yrt)-1.5))
max((max(yrt)+1.5))]); % set limits of axes
title('Returns for BHP, TLS, and AllOrds'); % add title
```



Q1(a) Exploratory Data Analysis

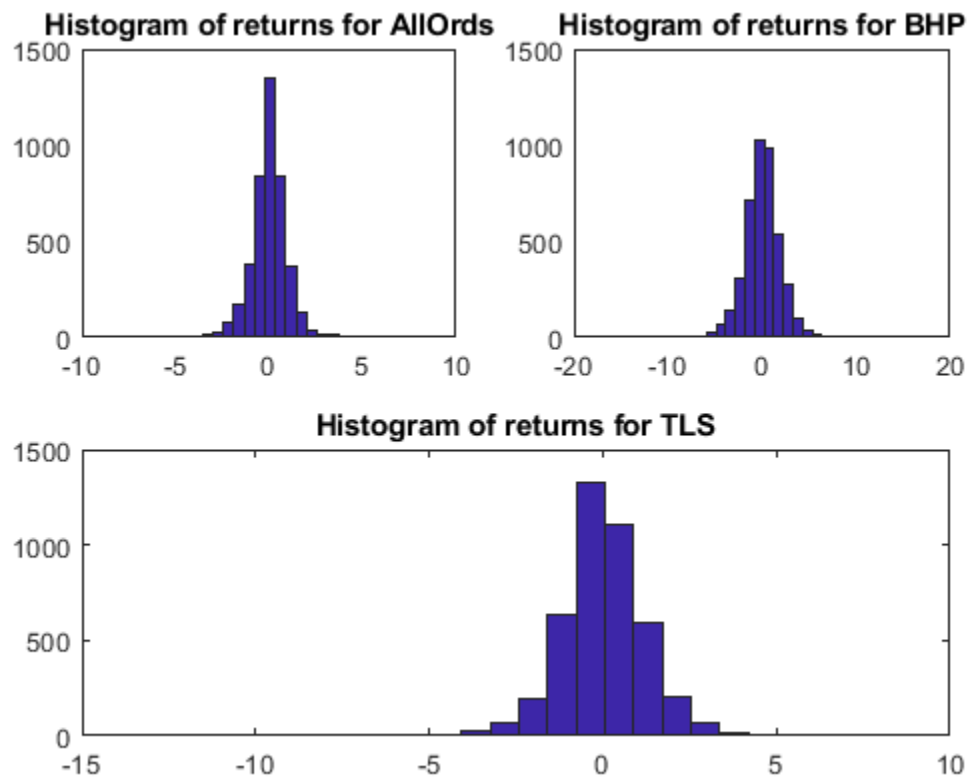
Descriptive Statistics

```
[mean(yrt); median(yrt); std(yrt); min(yrt); max(yrt); skewness(yrt);  
kurtosis(yrt)]
```

```
% Histograms of return series  
figure;subplot(2,2,1);hist(yrt(:,1),25);title('Histogram of returns  
for AllOrds');  
subplot(2,2,2);hist(yrt(:,2),25);title('Histogram of returns for  
BHP');  
subplot(2,1,2);hist(yrt(:,3),25);title('Histogram of returns for  
TLS');
```

ans =

0.0122	0.0397	0.0235
0.0484	0.0544	0
0.9757	1.9598	1.2668
-8.5536	-14.0772	-12.3546
5.3601	11.4645	8.3427
-0.5770	-0.1760	-0.5062
8.9446	6.5566	9.2237



Q1(b) Determine forecast models and provide initial forecasts

```
n=length(yrt); % total number of observation
ns=round(0.75*n); % approx 75% of observations
nf=n-ns; % remaining ~25% of observations
ret_f = yrt(ns+1:end,:); %forecast sample data is last 25% of days.
ret_is = yrt(1:ns,:); % in-sample is first 75% of days.

ret_f1 = ret_is(end,2:3); % 1st forecast vector is of
    last day's return
ret_f2 = mean(ret_is(end-21:end,2:3)); % 2nd forecast vector is mean
    of last 22 days (~ 1 month) of in-sample period
ret_f3 = mean(ret_is(end-249:end,2:3)); % 3rd forecast is mean of
    last 250 days (~ 1 year) of in-sample period
ret_f4 = zeros(1,2); %setting up space for 4th model, regression
ret_f5 = zeros(1,2); %setting up space for 5th, ARMA model forecasts
ret_f6 = zeros(1,2); %setting up space for 6th, Reg-ARMA forecasts
ret_f7 = zeros(1,2); %setting up space for 7th, Reg-AR(1,season)
    forecasts

% 4th model is regression of BHP with lagged market index
xmat=[ones(ns-1,1) ret_is(1:end-1,1)]; % creates X matrix for
    regression
[B1,BINT1,R1,RINT1,STATS1] = regress(ret_is(2:ns,2),xmat);
% Regression Coefficients
B1
% Regression R-Squared
STATS1(1)
% Regression SER
STATS1(4)

ret_f4(1) = [1 ret_is(end,1)]*B1; % predicted value of regression for
    last period of in-sample

%4th model is regression of TLS with lagged market
xmat=[ones(ns-1,1) ret_is(1:end-1,1)]; % creates X matrix for
    regression
[B2,BINT2,R2,RINT2,STATS2] = regress(ret_is(2:ns,3),xmat);
ret_f4(2) = [1 ret_is(end,1)]*B2; % predicted value of regression for
    last period of in-sample
% Regression Coefficients
B2
% Regression R-Squared
STATS2(1)
% Regression SER
STATS2(4)

%5th method forecasts are from a suitable ARMA model chosen for each
    series.
%BHP
figure;subplot(2,1,1);plot(datenum(ydat(:,1:ns)),ret_is(:,2));
```

```

datetick('x','mm/yy'); % sets format for axis labels and preserves
    ticks and limits and keeps them in this format
axis([min(datenum(ydat(:,1:ns))) max(datenum(ydat(:,1:ns)))
    (min(yrt(:,2))-1.5) (max(yrt(:,2))+1.5)]); % set limits of axes
title('In-sample BHP returns'); % add title
subplot(2,1,2);autocorr(ret_is(:,2), 25);

% LB test for AR effects on BHP
[H5, pval5, Qs5, CV5] = lbqtest(ret_is(:,2), 5, 0.05);
[H10, pval10, Qs10, CV10] = lbqtest(ret_is(:,2), 10, 0.05);
[pval5 pval10]

% I choose AR(3) for BHP, now fit AR(3) and forecast
Mdl=arima(3,0,0);% specifies the AR(3) model
[EstMdl,EstParamCov,logL,info] = estimate(Mdl,ret_is(:,2)); %
    estimates the AR(3) model
[ret_f5(1), FMSE] = forecast(EstMdl,1,'Y0',ret_is(:,2)); % 1-period
    forecasts

%TLS
figure;subplot(2,1,1);plot(datenum(ydat(:,1:ns)),ret_is(:,3));
datetick('x','mm/yy'); % sets format for axis labels and preserves
    ticks and limits and keeps them in this format
axis([min(datenum(ydat(:,1:ns))) max(datenum(ydat(:,1:ns)))
    (min(yrt(:,3))-1.5) (max(yrt(:,3))+1.5)]); % set limits of axes
title('In-sample TLS returns'); % add title
subplot(2,1,2);autocorr(ret_is(:,3), 25);

% LB test for AR effects on TLS
[H5, pval5, Qs5, CV5] = lbqtest(ret_is(:,3), 5, 0.05);
[H10, pval10, Qs10, CV10] = lbqtest(ret_is(:,3), 10, 0.05);
[pval5 pval10]

% I choose AR(3) for TLS too, now fit AR(3) and forecast
Mdl=arima(3,0,0);% specifies the AR(3) model
[EstMdl,EstParamCov,logL,info] = estimate(Mdl,ret_is(:,3)); %
    estimates the AR(3) model
[ret_f5(2), FMSE] = forecast(EstMdl,1,'Y0',ret_is(:,3)); % 1-period
    forecasts

%6th method forecasts are from a suitable Reg-ARMA model chosen for
    each series.
figure;subplot(2,1,1);plot(datenum(ydat(:,2:ns)),R1); % R1 are
    residuals from regression of BHP on lagged market
datetick('x','mm/yy'); % sets format for axis labels and preserves
    ticks and limits and keeps them in this format
axis([min(datenum(ydat(:,1:ns))) max(datenum(ydat(:,1:ns)))
    min(R1)-1.5 max(R1)+1.5]); % set limits of axes
title('Residuals from regression of BHP on lagged market'); % add
    title
subplot(2,1,2);autocorr(R1,25);
% LB test for AR effects on BHP regression residuals
[H5, pval5, Qs5, CV5] = lbqtest(R1, 5, 0.05,4);
[H10, pval10, Qs10, CV10] = lbqtest(R1, 10, 0.05,9);

```

```

[pval5 pval10]

% I choose Reg-AR(3) for BHP, now fit and forecast
Mdl=arima(3,0,0);% specifies the AR(3) model
[EstMdl,EstParamCov,logL,info] =
    estimate(Mdl,ret_is(4:end,2),'Y0',ret_is(1:3,2),'X',ret_is(3:end-1,1));%
    estimates the AR(3) model with lagged AORds as X variable
[ret_f6(1), FMSE] =
    forecast(EstMdl,1,'Y0',ret_is(4:end,2),'X0',ret_is(3:end-1,1),'XF',ret_is(end,1))
    1-period forecasts

%TLS
figure;subplot(2,1,1);plot(datenum(ydat(:,2:ns)),R2); % R2 are
    residuals from regression of BHP on lagged market
datetick('x','mm/yy'); % sets format for axis labels and preserves
    ticks and limits and keeps them in this format
axis([min(datenum(ydat(:,1:ns))) max(datenum(ydat(:,1:ns)))
    min(R2)-1.5 max(R2)+1.5]); % set limits of axes
title('Residuals from regression of TLS on lagged market'); % add
    title
subplot(2,1,2);autocorr(R2,25);

% LB test for AR effects on TLS regression residuals
[H5, pval5, Qs5, CV5] = lbqtest(R2, 5, 0.05,3);
[H10, pval10, Qs10, CV10] = lbqtest(R2, 10, 0.05,8);
[pval5 pval10]

% I choose Reg-AR(3) for TLS too, now fit AR(3) and forecast
Mdl=arima(3,0,0);% specifies the AR(3) model
[EstMdl,EstParamCov,logL,info] =
    estimate(Mdl,ret_is(4:end,3),'Y0',ret_is(1:3,3),'X',ret_is(3:end-1,1));%
    estimates the AR(3) model with lagged AORds as X variable
[ret_f6(2), FMSE] =
    forecast(EstMdl,1,'Y0',ret_is(4:end,3),'X0',ret_is(3:end-1,1),'XF',ret_is(end,1))
    1-period forecasts

%7th method forecasts are from a suitable Reg-ARMA plus seasonal model
    chosen for each series.
% I choose Reg-AR(5) plus 5th lag for BHP, now fit and forecast
Mdl=arima(5,0,0);% specifies the AR(5) model
[EstMdl,EstParamCov,logL,info] =
    estimate(Mdl,ret_is(6:end,2),'Y0',ret_is(1:5,2),'X',ret_is(5:end-1,1));%
    estimates the AR(5) model with lagged AORds as X variable
[ret_f7(1), FMSE] =
    forecast(EstMdl,1,'Y0',ret_is(6:end,2),'X0',ret_is(5:end-1,1),'XF',ret_is(end,1))
    1-period forecasts
%TLS
% I choose Reg-AR(5) for TLS too, now fit and forecast
Mdl=arima(5,0,0);% specifies the AR(5) model
[EstMdl,EstParamCov,logL,info] =
    estimate(Mdl,ret_is(6:end,3),'Y0',ret_is(1:5,3),'X',ret_is(5:end-1,1));%
    estimates the AR(5) model with lagged AORds as X variable

```

```
[ret_f7(2), FMSE] =
    forecast(EstMdl,1,'Y0',ret_is(6:end,3),'X0',ret_is(5:end-1,1),'XF',ret_is(end,1))
    1-period forecasts
```

```
% Note that these are all vectors, representing the forecasts for both
    series BHP and TLS.
% these forecasts should be compared with ret_f(1,:)
```

```
B1 =
```

```
    0.0572
   -0.1019
```

```
ans =
```

```
    0.0027
```

```
ans =
```

```
    4.0328
```

```
B2 =
```

```
    0.0144
   -0.0624
```

```
ans =
```

```
    0.0022
```

```
ans =
```

```
    1.8357
```

```
ans =
```

```
    0.0091    0.0211
```

```
ARIMA(3,0,0) Model:
```

```
-----
```

```
Conditional Probability Distribution: Gaussian
```

Parameter	Value	Standard Error	t Statistic
-----	-----	-----	-----
Constant	0.0615229	0.0357121	1.72275

AR{1}	-0.0407564	0.011693	-3.48553
AR{2}	-0.0202143	0.0113742	-1.77721
AR{3}	-0.0504812	0.0106235	-4.75185
Variance	4.02391	0.0633256	63.5432

ans =

1.0e-05 *

0.3202 0.9655

ARIMA(3,0,0) Model:

Conditional Probability Distribution: Gaussian

Parameter	Value	Standard Error	t Statistic
-----	-----	-----	-----
Constant	0.0143704	0.0244296	0.588237
AR{1}	0.0349069	0.0118865	2.93669
AR{2}	-0.0734759	0.0133049	-5.52247
AR{3}	-0.0544693	0.0142012	-3.83554
Variance	1.82095	0.0233849	77.8685

ans =

0.0488 0.0800

ARIMAX(3,0,0) Model:

Conditional Probability Distribution: Gaussian

Parameter	Value	Standard Error	t Statistic
-----	-----	-----	-----
Constant	0.0596788	0.035767	1.66854
AR{1}	-0.00911862	0.0178526	-0.510774
AR{2}	-0.019058	0.011357	-1.67809
AR{3}	-0.0482006	0.0106602	-4.52154
Beta1	-0.0849911	0.0336812	-2.5234
Variance	4.01533	0.0632663	63.4671

ans =

1.0e-06 *

0.0142 0.1299

ARIMAX(3,0,0) Model:

Conditional Probability Distribution: Gaussian

Parameter	Value	Standard Error	t Statistic
Constant	0.0163039	0.0244403	0.66709
AR{1}	0.063913	0.0134256	4.76054
AR{2}	-0.0756578	0.0133463	-5.66884
AR{3}	-0.0534658	0.0141896	-3.76797
Beta1	-0.0987733	0.0184123	-5.36452
Variance	1.81199	0.023426	77.3496

ARIMAX(5,0,0) Model:

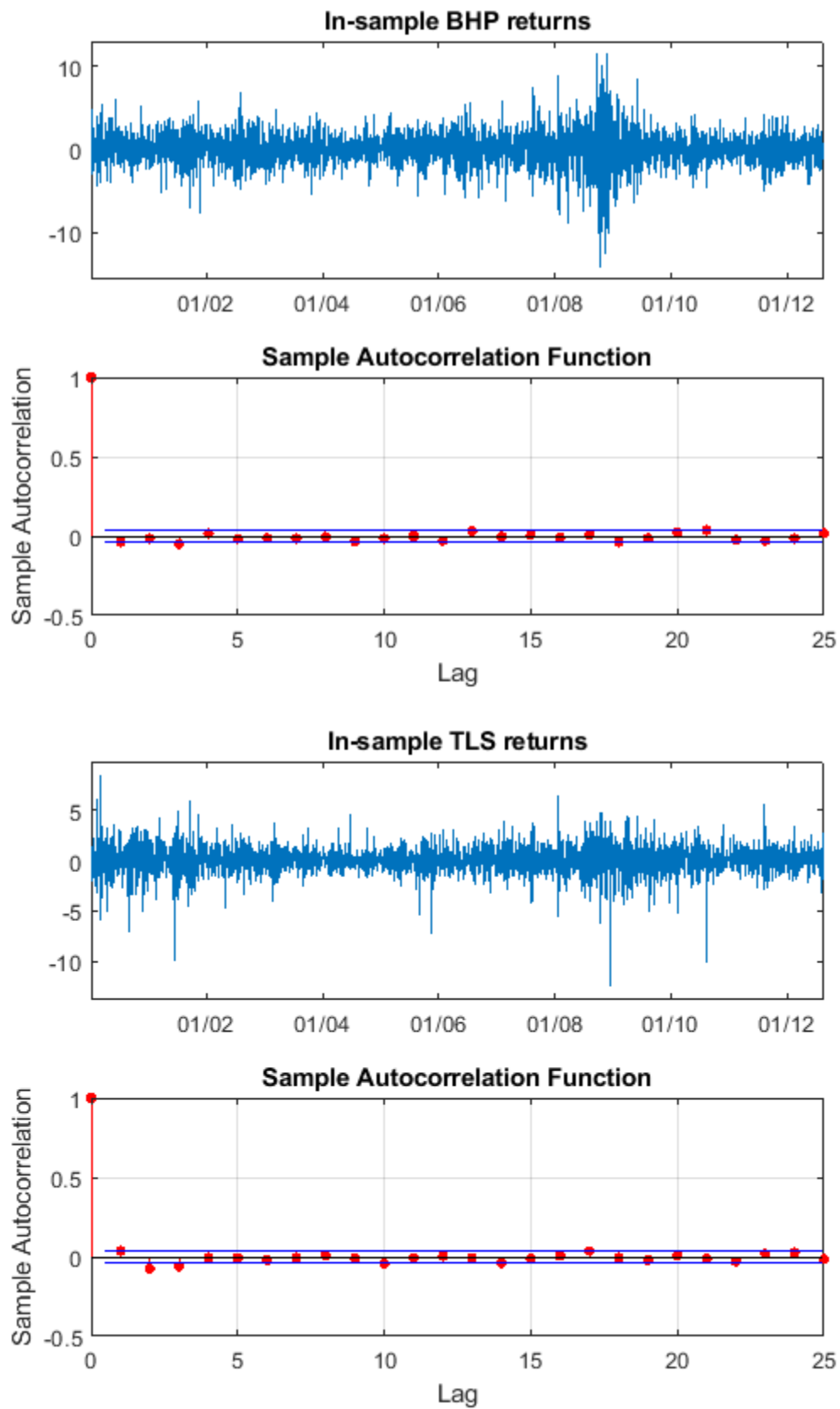
Conditional Probability Distribution: Gaussian

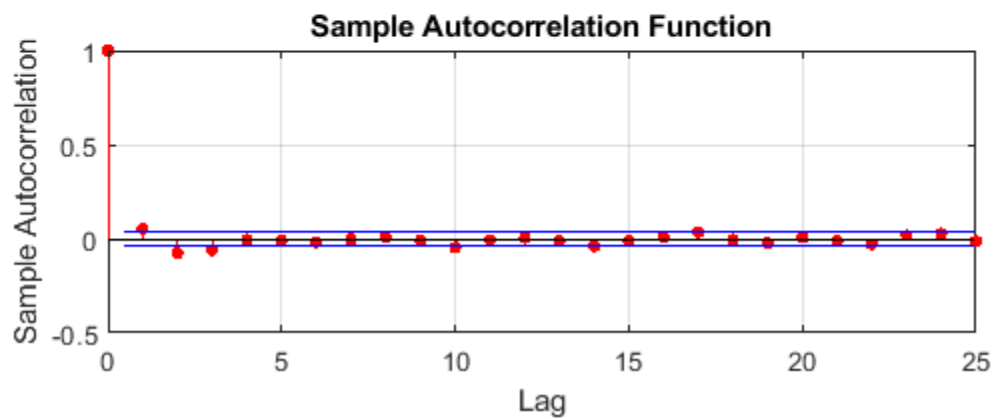
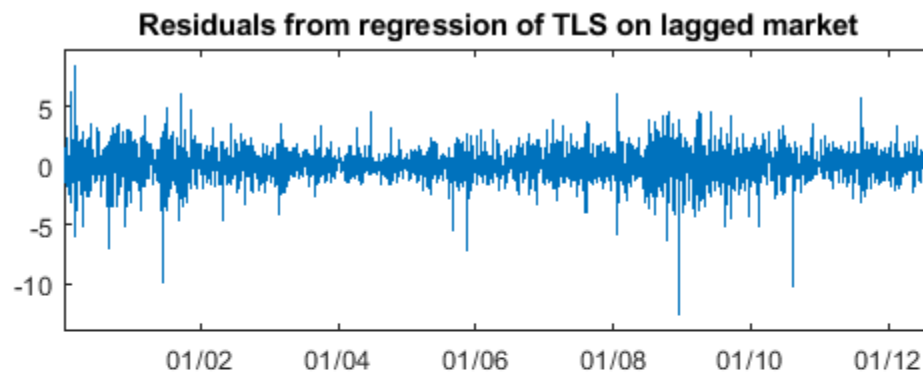
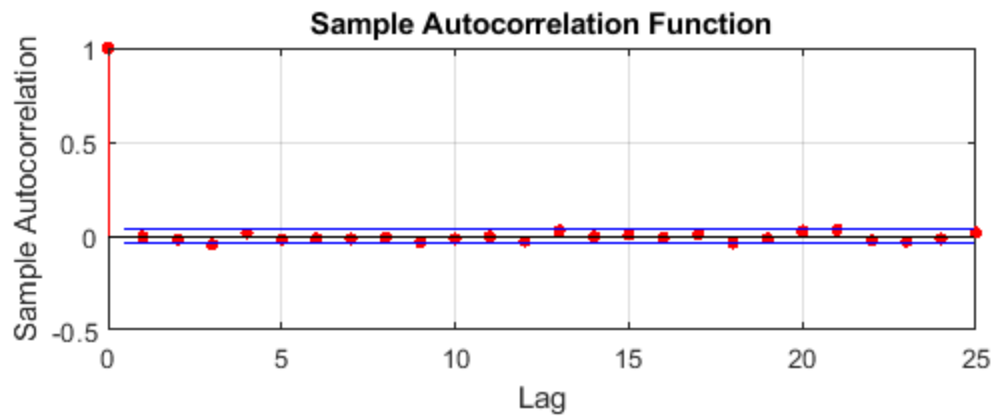
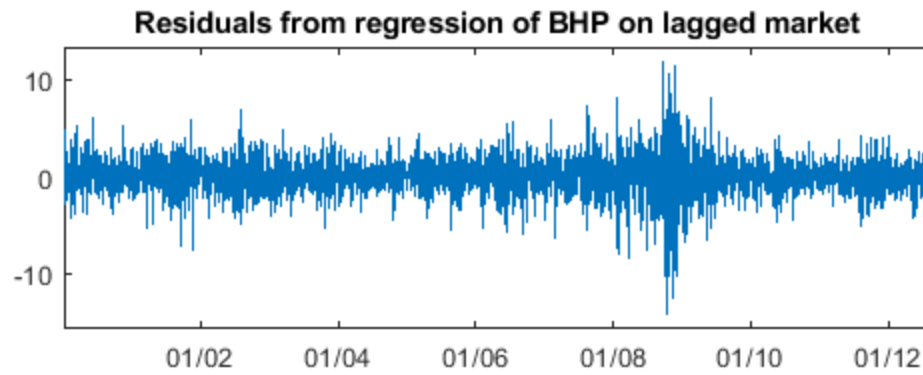
Parameter	Value	Standard Error	t Statistic
Constant	0.0606042	0.0358564	1.69019
AR{1}	-0.00940982	0.0179891	-0.523085
AR{2}	-0.0187566	0.0114001	-1.6453
AR{3}	-0.0478407	0.0106616	-4.4872
AR{4}	0.0115363	0.0122522	0.941567
AR{5}	-0.0180055	0.0106025	-1.69823
Beta1	-0.0821319	0.0338337	-2.42752
Variance	4.01351	0.0643006	62.4179

ARIMAX(5,0,0) Model:

Conditional Probability Distribution: Gaussian

Parameter	Value	Standard Error	t Statistic
Constant	0.0166396	0.0246172	0.675933
AR{1}	0.0642701	0.0134799	4.76784
AR{2}	-0.077287	0.0134276	-5.75585
AR{3}	-0.0540506	0.0142975	-3.78041
AR{4}	-0.00631422	0.0155509	-0.406035
AR{5}	-0.0161318	0.0154038	-1.04726
Beta1	-0.0991773	0.0184802	-5.36667
Variance	1.81132	0.0242641	74.6503





Q1(c) Moving origin forecasts

```
for t=2:nf

    ret_is = yrt(t:ns+t-1,:); % in-sample is all days before the last
    25% of days.
    ret_f1(t,:) = ret_is(end,2:3); % 1st forecast vector
    is of last day's return
    ret_f2(t,:) = mean(ret_is(end-21:end,2:3)); % 2nd forecast vector
    is mean of last 22 days (~ 1 month) of in-sample period
    ret_f3(t,:) = mean(ret_is(end-249:end,2:3)); % 3rd forecast is mean
    of last 250 days (~ 1 year) of in-sample period

%4th model is regression of BHP with lagged market
    xmat=[ones(ns-1,1) ret_is(1:end-1,1)]; % creates X matrix for
    regression
    [B1,BINT1,R1,RINT1,STATS1] = regress(ret_is(2:ns,2),xmat);
    ret_f4(t,1) = [1 ret_is(end,1)]*B1;
%4th model is regression of TLS with lagged market
    xmat=[ones(ns-1,1) ret_is(1:end-1,1)]; % creates X matrix for
    regression
    [B2,BINT2,R2,RINT2,STATS2] = regress(ret_is(2:ns,3),xmat);
    ret_f4(t,2) = [1 ret_is(end,1)]*B2;
%5th
Mdl=arima(3,0,0);[EstMdl,EstParamCov,logL,info] =
    estimate(Mdl,ret_is(:,2),'display','Off');% estimates the AR(3) model
[ret_f5(t,1), FMSE] = forecast(EstMdl,1,'Y0',ret_is(:,2));% 1-period
    forecasts
Mdl=arima(3,0,0);[EstMdl,EstParamCov,logL,info] =
    estimate(Mdl,ret_is(:,3), 'display','Off');% estimates the AR(3)
    model
[ret_f5(t,2), FMSE] = forecast(EstMdl,1,'Y0',ret_is(:,3));% 1-period
    forecasts
%6th
Mdl=arima(3,0,0);[EstMdl,EstParamCov,logL,info] =
    estimate(Mdl,ret_is(4:end,2),'Y0',ret_is(1:3,2),'X',ret_is(3:end-1,1),'display','
    estimates the AR(3) model with lagged AORds as X variable
[ret_f6(t,1), FMSE] =
    forecast(EstMdl,1,'Y0',ret_is(4:end,2),'X0',ret_is(3:end-1,1),'XF',ret_is(end,1))
    1-period forecasts
Mdl=arima(3,0,0);[EstMdl,EstParamCov,logL,info] =
    estimate(Mdl,ret_is(4:end,3),'Y0',ret_is(1:3,3),'X',ret_is(3:end-1,1),'display','
    estimates the AR(3) model with lagged AORds as X variable
[ret_f6(t,2), FMSE] =
    forecast(EstMdl,1,'Y0',ret_is(4:end,3),'X0',ret_is(3:end-1,1),'XF',ret_is(end,1))
    1-period forecasts
%7th
Mdl=arima(5,0,0);[EstMdl,EstParamCov,logL,info] =
    estimate(Mdl,ret_is(6:end,2),'Y0',ret_is(1:5,2),'X',ret_is(5:end-1,1),'display','
    estimates the AR(3) model with lagged AORds as X variable
[ret_f7(t,1), FMSE] =
    forecast(EstMdl,1,'Y0',ret_is(6:end,2),'X0',ret_is(5:end-1,1),'XF',ret_is(end,1))
    1-period forecasts
```

```

Mdl=arima(5,0,0);[EstMdl,EstParamCov,logL,info] =
    estimate(Mdl,ret_is(6:end,3),'Y0',ret_is(1:5,3),'X',ret_is(5:end-1,1),'display','
    estimates the AR(3) model with lagged AORDs as X variable
[ret_f7(t,2), FMSE] =
    forecast(EstMdl,1,'Y0',ret_is(6:end,3),'X0',ret_is(5:end-1,1),'XF',ret_is(end,1))
    1-period forecasts
end

% we now have 1068 days of one-step-ahead forecasts for both asset
    series using each of 7 different methods.

% Assess forecast accuracy for all methods and for both series
%BHP
figure;plot(datenum(ydat(:,ns+1:end)),ret_f(:,2),'g'); % include dates
    and change order so AOrd is plotted last and is then on top
hold on;plot(datenum(ydat(:,ns+1:end)),ret_f1(:,1),'k*');
plot(datenum(ydat(:,ns+1:end)),ret_f2(:,1),'rd');
plot(datenum(ydat(:,ns+1:end)),ret_f3(:,1),'m^');
plot(datenum(ydat(:,ns+1:end)),ret_f4(:,1),'ko');
plot(datenum(ydat(:,ns+1:end)),ret_f5(:,1),'m+');
plot(datenum(ydat(:,ns+1:end)),ret_f6(:,1),'bp');
plot(datenum(ydat(:,ns+1:end)),ret_f7(:,1),'cd');
datetick('x','mm/yy','kepticks','keeplimits'); % sets format for axis
    labels and preserves ticks and limits and keeps them in this format
title('Actual returns and forecasts for BHP'); % add title
axis([min(datenum(ydat(:,ns+1:end))) max(datenum(ydat(:,ns+1:end)))
    min(ret_f(:,2))-0.5 max(ret_f(:,2))+0.5]);

[rmse1, mad1, map1]=getfa(ret_f(:,2),ret_f1(:,1));[rmse2, mad2,
    map2]=getfa(ret_f(:,2),ret_f2(:,1));
[rmse3, mad3, map3]=getfa(ret_f(:,2),ret_f3(:,1));[rmse4, mad4,
    map4]=getfa(ret_f(:,2),ret_f4(:,1));
[rmse5, mad5, map5]=getfa(ret_f(:,2),ret_f5(:,1));[rmse6, mad6,
    map6]=getfa(ret_f(:,2),ret_f6(:,1));
[rmse7, mad7, map7]=getfa(ret_f(:,2),ret_f7(:,1));
[rmse1 rmse2 rmse3 rmse4 rmse5 rmse6 rmse7;mad1 mad2 mad3 mad4 mad5
    mad6 mad7]

%TLS
figure;plot(datenum(ydat(:,ns+1:end)),ret_f(:,3),'g'); % include dates
    and change order so AOrd is plotted last and is then on top
hold on;plot(datenum(ydat(:,ns+1:end)),ret_f1(:,2),'k*');
plot(datenum(ydat(:,ns+1:end)),ret_f2(:,2),'rd');
plot(datenum(ydat(:,ns+1:end)),ret_f3(:,2),'m^');
plot(datenum(ydat(:,ns+1:end)),ret_f4(:,2),'ko');
plot(datenum(ydat(:,ns+1:end)),ret_f5(:,2),'m+');
plot(datenum(ydat(:,ns+1:end)),ret_f6(:,2),'bp');
plot(datenum(ydat(:,ns+1:end)),ret_f7(:,2),'cd');
datetick('x','mm/yy','kepticks','keeplimits'); % sets format for axis
    labels and preserves ticks and limits and keeps them in this format
title('Actual returns and forecasts for TLS'); % add title
axis([min(datenum(ydat(:,ns+1:end))) max(datenum(ydat(:,ns+1:end)))
    min(ret_f(:,3))-0.5 max(ret_f(:,3))+0.5]);

```

```
[rmse1, mad1, map1]=getfa(ret_f(:,3),ret_f1(:,2));[rmse2, mad2,
map2]=getfa(ret_f(:,3),ret_f2(:,2));
[rmse3, mad3, map3]=getfa(ret_f(:,3),ret_f3(:,2));[rmse4, mad4,
map4]=getfa(ret_f(:,3),ret_f4(:,2));
[rmse5, mad5, map5]=getfa(ret_f(:,3),ret_f5(:,2));[rmse6, mad6,
map6]=getfa(ret_f(:,3),ret_f6(:,2));
[rmse7, mad7, map7]=getfa(ret_f(:,3),ret_f7(:,2));
[rmse1 rmse2 rmse3 rmse4 rmse5 rmse6 rmse7;mad1 mad2 mad3 mad4 mad5
mad6 mad7]
```

```
% Note: the getfa.m cannot compute MAPE here as a number of daily
returns are zero because
% stock closes at same price as previous day (even though it may have
moved around during the day),
% which results in the APE dividing by zero and is undefined
```

```
ans =
```

```
'y=0, no mape'
```

```
ans =
```

```
'y=0, no mape'
```

```
ans =
```

```
'y=0, no mape'
```

```
ans =
```

```
'y=0, no mape'
```

```
ans =
```

```
'y=0, no mape'
```

```
ans =
```

```
'y=0, no mape'
```

```
ans =
```

```
'y=0, no mape'
```

```
ans =
```

2.4618	1.8411	1.8021	1.8023	1.8046	1.8031	1.8041
1.8177	1.3419	1.3159	1.3180	1.3209	1.3195	1.3203

ans =

'y=0, no mape'

ans =

'y=0, no mape'

ans =

'y=0, no mape'

ans =

'y=0, no mape'

ans =

'y=0, no mape'

ans =

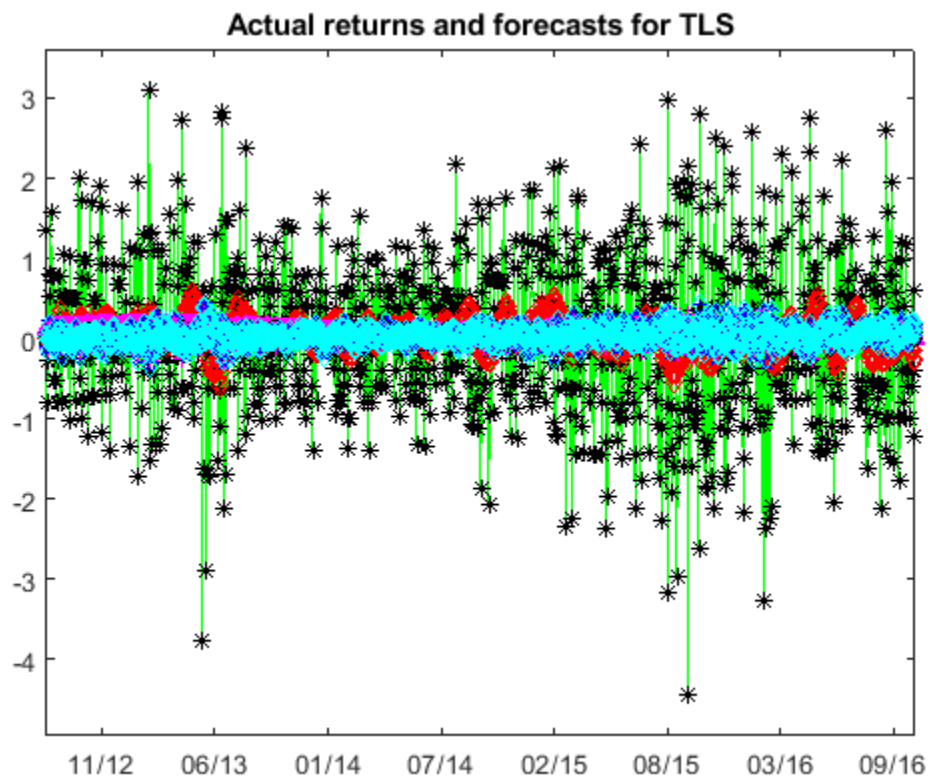
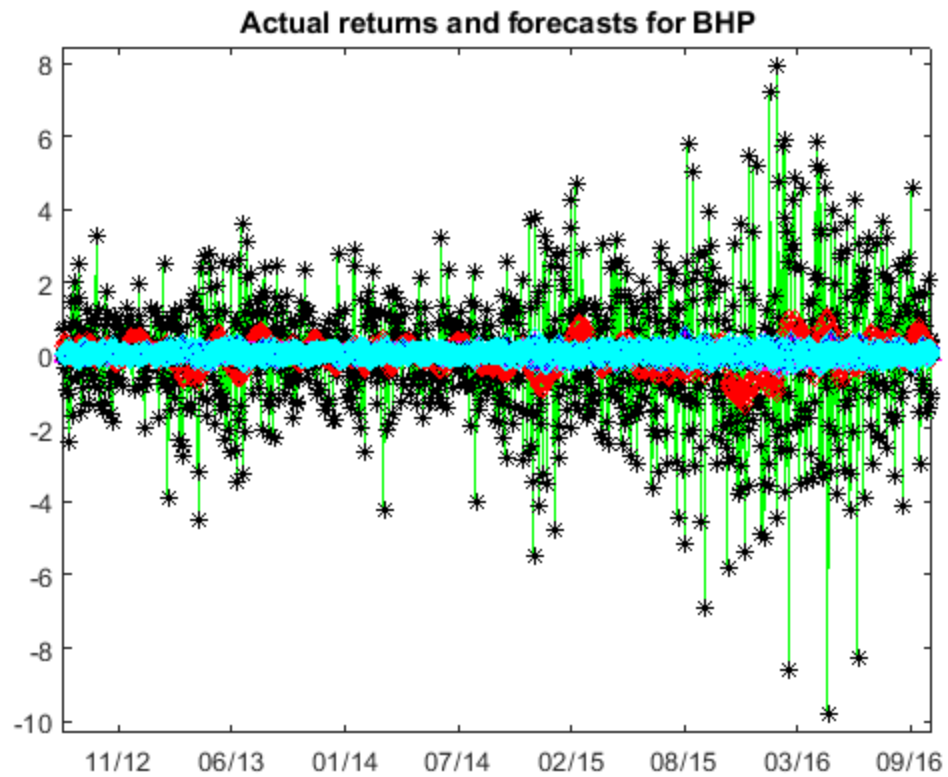
'y=0, no mape'

ans =

'y=0, no mape'

ans =

1.3664	0.9712	0.9492	0.9488	0.9497	0.9504	0.9512
1.0360	0.7510	0.7346	0.7338	0.7330	0.7341	0.7354



Published with MATLAB® R2017b