9.6/10

Ziming Huang — Econ 672 — October 25, 2019

Project 6

Exercise 1-Forecasting Variance

(A) The code of our parameter estimate function is as follow:

```
function [beta] = OLS(x,y)
%J is the number of independent variables(step number)
%y is the denpendent variable vector: J*1 t-series data
%x is the indenpendent variable matrix: J*n t-series data * n variables
% beta is the vestor of parameter: (n+1)*1
x(:,end+1)=1; % adding interception into regression
%beta=x'*y\(x'*x);
beta=(x'*x)\(x'*y);
end
```

The input of this function is independent variable data matrix $X = [X1, X2, \dots, X_n]$ and the dependent variable vector Y. The output is a vector of parameter $\hat{\beta} \equiv [\beta_1, \beta_2, \dots, \beta_n, \beta_0]$. Here β_0 is the parameter of interception.

(B) The function input is $[X, Y, J, T, step]$, where $X = [X1, X2, X3, \dots, X_n]$ and Y is the dependent variable vector, T is the latest data we used for estimating, J is the row number of independent variable data and $step$ is the order of model.

If we set estimate window length is 1000, then J=999, $T = 1000, 1001, \dots, T_{All} - 1)$. For AR1 and HAR1 model, we have $step = 1$.

Here is the table of forecast value and MSE from different models.

Table 1: Model Forecast value and MSE (with W=1000)

Stock	Real $RV_{T_{2769}}$	AR1	HAR1	Non-change
PG	4.2813×10^{-5}	4.7814×10^{-5}	4.2813×10^{-5}	1.3228×10^{-5}
DIS	3.1136×10^{-5}	5.6526×10^{-5}	6.7252×10^{-5}	2.9385×10^{-5}
		MSE_{AR1}	MSE_{HAR1}	MSE_{NC}
PG		1.2432×10^{-8}	1.1566×10^{-8}	1.6612×10^{-8}
DIS		1.1391×10^{-8}	9.8089×10^{-9}	1.2085×10^{-8}

From this table,for both stock, the range of MSE of these three model for PG and DIS is (1.1566×10^{-8},1.6612×10^{-8}) and (9.8089×10^{-9},1.2085×10^{-8}), respectively. All three models' MSE is less than 2×10^{-8} and they do well in the forecasting of RV. Among these models, HAR1 model has the smallest MSE and Non-change Model has the largest MSE. HAR1 does best according to MSE criterion.

AR Model

```
function [fc,beta] = AR(X,Y,J,T,step)
%J is the width of data window
%T stop day
```

```
%T+1 is forecasting day(newest available data-1)
%n is forecasting step number
%Y and X are available total data
x=zeros(J,step);
y1=zeros(step,1);
  for i=1:step
    x(:,i)=X(T-J+1-i:T-i);
    y1(i)=X(T-i+1);
  end
%y is the denpendent variable vector: J*1 t-series data
%x is the indenpendent variable matrix: J*n t-series data * n variables
%beta is the vestor of parameter: (n+1)*1
x=flipud(x);
y=flipud(Y(T-J+1:T));
beta=OLS(x,y);
%forecasting
y1=[y1;1];
fc=beta'*y1;
if fc<min(X(T-J+1-step:T)) || fc>max(X(T-J+1-step:T))
     fc=mean(X(T-J+1-step:T));
end

end
```

HAR Model

```
function [fc,beta] = HAR(X,Y,J,T,step)
%J is the width of X data window
%T+1 is forecasting day(newest available data-1)
%n is forecasting step number
%Y and X are available total data
x=zeros(J,3*step);
y1=zeros(3*step,1);
  for i=1:step
    x(:,3*i-2)=X(T-J+1-i:T-i);
    x(:,3*i-1)=movmean(X(T-J+1-i:T-i),[4,0]);
    x(:,3*i)=movmean(X(T-J+1-i:T-i),[21,0]);
    y1(3*i-2)=X(T-i+1);
    y1(3*i-1)=mean(X(T-i+1-4:T-i+1));
    y1(3*i)=mean(X(T-i+1-21:T-i+1));
  end
%y is the denpendent variable vector: J*1 t-series data
%x is the indenpendent variable matrix: J*n t-series data * n variables
%beta is the vestor of parameter: (n+1)*1
x=flipud(x(22:end,:));
```

```
y=flipud(Y(T-J+1+21:T));
beta=OLS(x,y);
  %forecasting
  y1=[y1;1];
  fc=beta'*y1;

if fc<min(X(T-J+1-step:T)) || fc>max(X(T-J+1-step:T))
     fc=mean(X(T-J+1-step:T));
end



end
```

(C) Here is the table of forecasting MSE from different forecasting models.

Table 2: Tabel of Model Forecast MSE

Stock	Window Size	MSE_{AR1}	MSE_{HAR1}	MSE_{NC}
PG	W=250	3.2466×10^{-8}	3.3395×10^{-8}	4.5041×10^{-8}
	W=500	1.7451×10^{-8}	1.6381×10^{-8}	1.9376×10^{-8}
DIS	W=250	8.6868×10^{-8}	1.0468×10^{-7}	1.2226×10^{-7}
	W=500	2.9928×10^{-8}	2.4585×10^{-8}	3.3995×10^{-8}

From this table, the range of models' MSE for PG and DIS when W=250 and is (3.2466×10^{-8},4.5041×10^{-8}) and (8.6868×10^{-8},1.2226×10^{-7}), respectively; the range of models' MSE for PG and DIS when W=500 and is (1.6381×10^{-8},1.9376×10^{-8}) and (2.4585×10^{-8},3.3995×10^{-8}), respectively.

Combine the results in part (B), We can find as the window size decreases from 1000 t0 250, the MSE of each model increases. There is no model consistently better than others when evaluated using different window size.

(D) Here is the plot of model MSE with different window size.

(+0.3) nice formatted plot

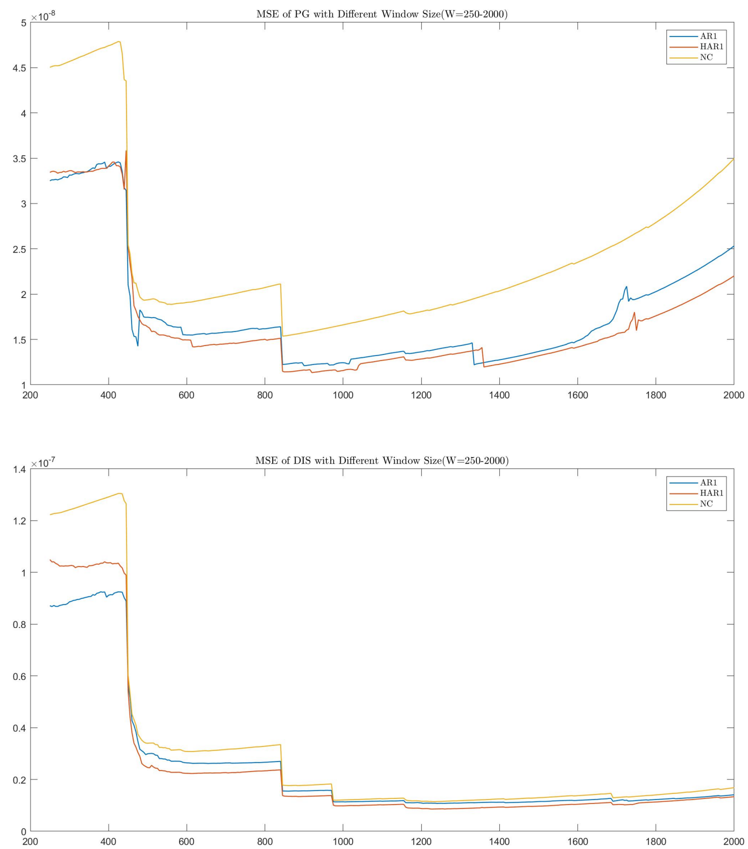


Figure 1: MSE with Different Window Size

From the figure we can see, the MSE of forecast for these three models show a decrease jump at around J=500 and reach the lowest value at around J=1000. After reaching the lowest point, the MSE of PG starts increase, however, the MSE of DIS is persistent for increasing J.

The **MATLAB** code:

Function of Local Variance

Script of Q1

```
addpath('D:\ZM-Documents\MATLAB\data\Stocks5Min','functions','scripts');
[dates_PG,lp_PG]=load_stock('PG.csv','m');
N_PG=sum(floor(dates_PG(1,1))==floor(dates_PG(:,1)));% number of observations
    per day
T_PG=size(dates_PG,1)/N_PG;
[rdates_PG,lr_PG]=log_return([dates_PG lp_PG],N_PG,1);
```

```
days_PG=unique(floor(rdates_PG));

n=N_PG-1;
a=5;
[lr_c_PG,lr_d_PG]=c_d_log_returns(lr_PG,N_PG,a);

RV_PG=transpose(sum(lr_PG.^2));
QIV_PG=transpose(sum((n/3)*(lr_c_PG.^4)));

%1B
step=1;
X=RV_PG;
Y=RV_PG;
J=1000-1;%row number of independent variable
fRV_AR1_PG1=zeros(T_PG-J-1,1);%t increase
fRV_HAR1_PG1=zeros(T_PG-J-1,1);
%fRV_NC_PG=zeros(T_PG-J);



for T=J+1:T_PG-1% the newest data used to estimate beta
    fRV_AR1_PG1(T-J)=AR(X,Y,J,T,step);
    fRV_HAR1_PG1(T-J)=HAR(X,Y,J,T,step);
end
ERR_AR1_PG1=(mean((Y(J+2:T_PG)-fRV_AR1_PG1).^2));
ERR_HAR1_PG1=(mean((Y(J+2:T_PG)-fRV_HAR1_PG1).^2));
ERR_CN_PG1=(mean((Y(J+2:T_PG)-Y(J+1:T_PG-1)).^2));


%1C
J=250-1;
fRV_AR1_PG2=zeros(T_PG-J-1,1);%t increase
fRV_HAR1_PG2=zeros(T_PG-J-1,1);
%fRV_NC_PG=zeros(T_PG-J);

for T=J+1:T_PG-1
    fRV_AR1_PG2(T-J)=AR(X,Y,J,T,step);
    fRV_HAR1_PG2(T-J)=HAR(X,Y,J,T,step);
end
ERR_AR1_PG2=(mean((Y(J+2:T_PG)-fRV_AR1_PG2).^2));
ERR_HAR1_PG2=(mean((Y(J+2:T_PG)-fRV_HAR1_PG2).^2));
ERR_CN_PG2=(mean((Y(J+2:T_PG)-Y(J+1:T_PG-1)).^2));
```

```matlab
J=500-1;
fRV_AR1_PG3=zeros(T_PG-J-1,1);%t increase
fRV_HAR1_PG3=zeros(T_PG-J-1,1);
fRV_NC_PG3=zeros(T_PG-J);

for T=J+1:T_PG-1
    fRV_AR1_PG3(T-J)=AR(X,Y,J,T,step);
    fRV_HAR1_PG3(T-J)=HAR(X,Y,J,T,step);
end
ERR_AR1_PG3=(mean((Y(J+2:T_PG)-fRV_AR1_PG3).^2));
ERR_HAR1_PG3=(mean((Y(J+2:T_PG)-fRV_HAR1_PG3).^2));
ERR_CN_PG3=(mean((Y(J+2:T_PG)-Y(J+1:T_PG-1)).^2));

%1D

% ERR_AR1_PG4=zeros(351,1);
% ERR_HAR1_PG4=zeros(351,1);
% ERR_CN_PG4=zeros(351,1);
% for J=250-1:5:2000-1
% fRV_AR1_PG1=zeros(T_PG-J-1,1);%t increase
% fRV_HAR1_PG1=zeros(T_PG-J-1,1);
%   for T=J+1:T_PG-1% the newest data used to estimate beta
%     fRV_AR1_PG1(T-J)=AR(X,Y,J,T,step);
%     fRV_HAR1_PG1(T-J)=HAR(X,Y,J,T,step);
%   end
% ERR_AR1_PG4((J+1)/5-49)=(mean((Y(J+2:T_PG)-fRV_AR1_PG1).^2));
% ERR_HAR1_PG4((J+1)/5-49)=(mean((Y(J+2:T_PG)-fRV_HAR1_PG1).^2));
% ERR_CN_PG4((J+1)/5-49)=(mean((Y(J+2:T_PG)-Y(J+1:T_PG-1)).^2));
% end
%
%
% figure;
% plot(250:5:2000,ERR_AR1_PG4,'linewidth',1);
% hold on;
% plot(250:5:2000,ERR_HAR1_PG4,'linewidth',1);
% hold on;
% plot(250:5:2000,ERR_CN_PG4,'linewidth',1);
% legend('AR1','HAR1','NC');
% title('MSE of PG with Different Window Size(W=250-2000)');
```

Exercise 2-Errors in Variables (EIV)

(A) The **MATLAB** code:

```
function [X,Y] = simulation(var1,var2,var3,N,beta)
noise=random('normal',0,sqrt(var3),N,1);
X=random('normal',0,sqrt(var1),N,1)+noise;
u=random('normal',0,sqrt(var2),N,1);
Y=beta'.*X+u;
end
```

In this function, $var1$ is the variance of X; $var2$ is the variance of error term u; $var3$ is the variance of $noise$. If $var3 = 0$, then the noise in X is zero.

(B) Here is the summary table of beta.

Table 3: Summary Table of Parameter

Parameter	β_0	β_1
Value	0.0376	1.0040

(C) Here are the histogram figures of β_0 and β_1.

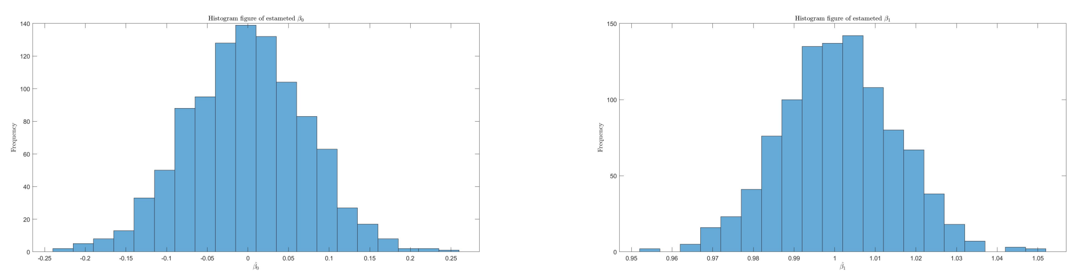
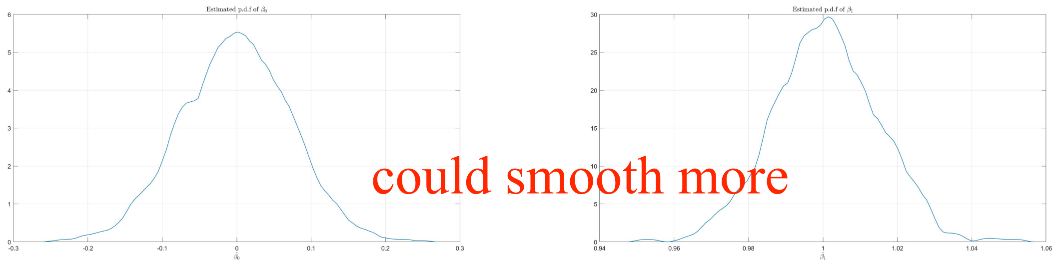


Figure 2: Histogram of β_0 and β_1

From the histogram figure of β_0 and β_1, we can see the shape of distribution of β_0 and β_1 is very similar to normal distribution.

Specifically, the mean of β_0 showed in the plot is zero and the mean of β_1 is 1. The result satisfies our expectation.

(D) Here are the figures of estimated probability density function of β_0 and β_1.



could smooth more

Figure 3: Estimated probability density function of β_0 and β_1

From the figures we can see the p.d.f. of β_0 and β_1 are very similar to normal distribution with mean equals to 0 and 1, respectively.

Compare the distribution shape of β_0 and β_1 we can find: β_0's distribution with a fatter tail and less peak, which indicates the variance of β_0 is larger than the variance of β_1.

(E) For regression:

$$\tilde{Y}_i = \tilde{X}_i \beta + \tilde{u}_i$$

we can estimate β by calculate:

$$\hat{\beta}^{ture} = \frac{Cov(\tilde{Y}_i, \tilde{X}_i)}{Var(\tilde{X}_i)} = \frac{Cov(\tilde{X}_i\beta, \tilde{X}_i) + Cov(\tilde{u}_i, \tilde{X}_i)}{Var(\tilde{X}_i)}$$

If some noises are added into $\tilde{X}_i$, the regression with noise will be:

$$\tilde{Y}_i = \tilde{X}_i^* \beta^* + \tilde{u}_i$$

where $\tilde{X}_i^* = \tilde{X}_i + \eta_i$ is the noise version of $\tilde{X}_i$. we can estimate β^* by calculate:

$$\hat{\beta}^* = \frac{Cov(\tilde{Y}_i, \tilde{X}_i^*)}{Var(\tilde{X}_i^*)} = \frac{Cov(\tilde{X}_i\beta, \tilde{X}_i) + Cov(\tilde{u}_i, \tilde{X}_i) + Cov(\tilde{X}_i\beta, \eta_i) + Cov(\tilde{u}_i, \eta_i)}{Var(\tilde{X}_i) + Var(\eta) + Cov(\tilde{X}_i, \eta_i)}$$

Assume noises η_i are uncorrelated with $\tilde{X}_i$, we can simplify $\hat{\beta}^*$ as:

$$\hat{\beta}^* = \frac{Cov(\tilde{X}_i\beta, \tilde{X}_i) + Cov(\tilde{u}_i, \tilde{X}_i)}{Var(\tilde{X}_i) + Var(\eta)}$$

as the $Var(\eta)$ increases, $\hat{\beta}^*$ will decrease and we will get a down-biased estimator of β. For the rest part, noises will be added to X to see how it affect the estimation of parameter β.

We first change the variance of *noise* from 0 to $0.30\sigma_x^2$.

(i) Here is the summary table of estimate value of $\hat{\beta}_0{}^*$ and $\hat{\beta}_1{}^*$.

Table 4: Summary Table of Parameter

Parameter	$\hat{\beta}_0{}^*$	$\hat{\beta}_1{}^*$
Value	0.1643	0.9881

(ii) Here are the histogram figures of $\hat{\beta}_0{}^*$ and $\hat{\beta}_1{}^*$.

density should changed after adding noise and beta goes
to around 0.75 0.65 respectively, check your results again (-1.5)

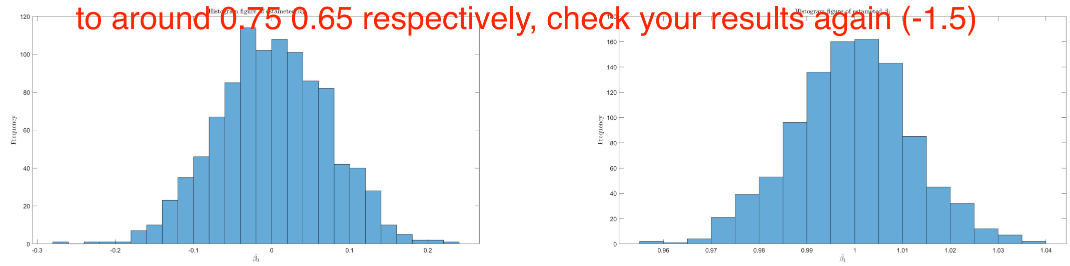


Figure 4: Histogram of $\hat{\beta}_0{}^*$ and $\hat{\beta}_1{}^*$

(iii) Here are the figures of estimated probability density function of $\hat{\beta}_0{}^*$ and $\hat{\beta}_1{}^*$.

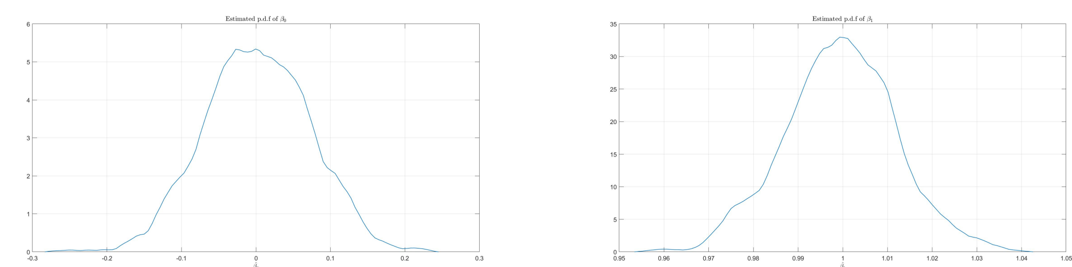


Figure 5: Estimated probability density function of $\hat{\beta}_0{}^*$ and $\hat{\beta}_1{}^*$

From the figures we can see the shape of distribution of $\hat{\beta}_0{}^*$ and $\hat{\beta}_1{}^*$ are still similar to normal distribution, but they show a trend of left skew and the mean of $\hat{\beta}_0{}^*$ and $\hat{\beta}_1{}^*$ in the estimated distribution are a little smaller than 0 and 1, respectively.

(F) We then change the variance of η_i to $0.50\sigma_x^2$.

(i) Here is the summary table of estimate value of $\hat{\beta_0}^*$ and $\hat{\beta_1}^*$.

Table 5: Summary Table of Parameter

Parameter	$\hat{\beta_0}^*$	$\hat{\beta_1}^*$
Value	0.0338	0.9835

(ii) Here are the histogram figures of $\hat{\beta_0}^*$ and $\hat{\beta_1}^*$.

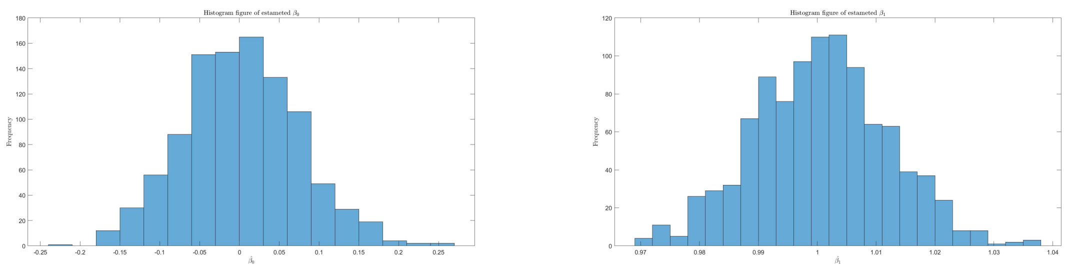


Figure 6: Histogram of $\hat{\beta_0}^*$ and $\hat{\beta_1}^*$

(iii) Here are the figures of estimated probability density function of $\hat{\beta_0}^*$ and $\hat{\beta_1}^*$.

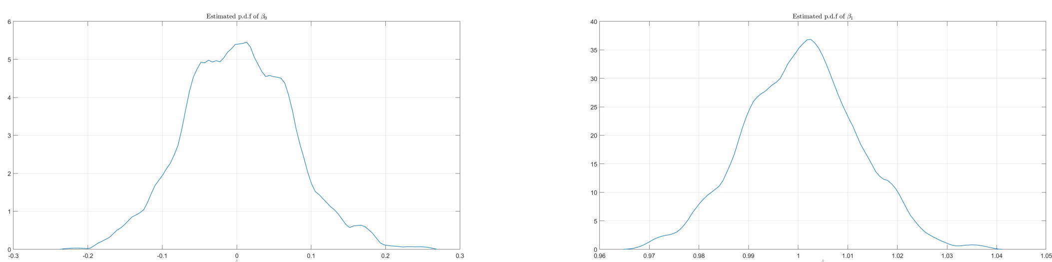


Figure 7: Histogram of $\hat{\beta_0}^*$ and $\hat{\beta_1}^*$

From the figures we can see the shape of distribution of $\hat{\beta_0}^*$ and $\hat{\beta_1}^*$ are little different to normal distribution: they are skewed left and the mean of $\hat{\beta_0}^*$ and $\hat{\beta_1}^*$ in the estimated distribution are smaller than $\hat{\beta_0}^*$ and $\hat{\beta_1}^*$ in part E.

If we keep increasing the variance of noises η_i, the distribution of $\hat{\beta_0}^*$ and $\hat{\beta_1}^*$ will skew much more to the left , and their mean will deviation more from true value of β_0 and β_1.

The **MATLAB**:

Scripts of Q2

```
addpath('D:\ZM—Documents\MATLAB\data','functions','scripts');


%2A
var1=25.2;
var2=0.5;% noise in Y
var3=0; % noise in X
beta=1;
N=100;
[X,Y]=simulation(var1,var2,var3,N,beta);
%2B
beta_hat=OLS(X,Y);
%3C
r=1000;
X1=zeros(N,r);
Y1=zeros(N,r);
beta_hat1=zeros(r,2);
for i=1:r
 [X1(:,i),Y1(:,i)]=simulation(var1,var2,var3,N,beta);
 beta_hat1(i,:)=OLS(X1(:,i),Y1(:,i));
end

figure;
histogram(beta_hat1(:,2),20);
xlabel('$\hat{\beta_0}$');
ylabel('Frequency');
title('Histogram figure of estameted $\beta_0$');
figure;
histogram(beta_hat1(:,1),20);
xlabel('$\hat{\beta_1}$');
ylabel('Frequency');
title('Histogram figure of estameted $\beta_1$');

%2D
[f_beta0,x_beta0]=ksdensity(beta_hat1(:,2),'Kernel','epanechnikov','Bandwidth
    ',0.01);
figure;
plot(x_beta0,f_beta0);
xlabel('$\hat{\beta_0}$');
grid on;
title('Estimated p.d.f of $\beta_0$');

[f_beta1,x_beta1]=ksdensity(beta_hat1(:,1),'Kernel','epanechnikov','Bandwidth
    ',0.002);
```

```
figure;
plot(x_beta1,f_beta1);
xlabel('$\hat{\beta_1}$');
grid on;
title('Estimated p.d.f of $\beta_1$');


%2E
%repeat A
var1=25.2;
var2=0.5;
var3=0.3*var1;
beta=1;
N=100;
[X_hat,Y_hat]=simulation(var1,var2,var3,N,beta);
%repeat B
%estimate beta_hat
beta_hat_1=OLS(X_hat,Y_hat);
%repeat C
%estimate different beta_hat
r=1000;
X2=zeros(N,r);
Y2=zeros(N,r);
beta_hat2=zeros(r,2);
for i=1:r
 [X2(:,i),Y2(:,i)]=simulation(var1,var2,var3,N,beta);
 beta_hat2(i,:)=OLS(X2(:,i),Y2(:,i));
end


figure;
histogram(beta_hat2(:,2));
xlabel('$\hat{\beta_0}$');
ylabel('Frequency');
title('Histogram figure of estameted $\beta_0$');
figure;
histogram(beta_hat2(:,1));
xlabel('$\hat{\beta_1}$');
ylabel('Frequency');
title('Histogram figure of estameted $\beta_1$');

%repeat C
%estimate distribution of beta_hat
```

```
[f_beta0,x_beta0]=ksdensity(beta_hat2(:,2),'Kernel','epanechnikov','Bandwidth
    ',0.01);
figure;
plot(x_beta0,f_beta0);
xlabel('$\hat{\beta_0}$');
grid on;
title('Estimated p.d.f of $\beta_0$');

[f_beta1,x_beta1]=ksdensity(beta_hat2(:,1),'Kernel','epanechnikov','Bandwidth
    ',0.002);
figure;
plot(x_beta1,f_beta1);
xlabel('$\hat{\beta_1}$');
grid on;
title('Estimated p.d.f of $\beta_1$');

%2F
%repeat A
var1=25.2;
var2=0.5;
var3=0.5*var1;
beta=1;
N=100;
[X_hat,Y_hat]=simulation(var1,var2,var3,N,beta);
%repeat B
%estimate beta_hat
beta_hat_3=OLS(X_hat,Y_hat);
%repeat C
%estimate different beta_hat
r=1000;
X2=zeros(N,r);
Y2=zeros(N,r);
beta_hat4=zeros(r,2);
for i=1:r
 [X2(:,i),Y2(:,i)]=simulation(var1,var2,var3,N,beta);
 beta_hat4(i,:)=OLS(X2(:,i),Y2(:,i));
end


figure;
histogram(beta_hat4(:,2));
xlabel('$\hat{\beta_0}$');
ylabel('Frequency');
title('Histogram figure of estameted $\beta_0$');
```

```
figure;
histogram(beta_hat4(:,1));
xlabel('$\hat{\beta_1}$');
ylabel('Frequency');
title('Histogram figure of estameted $\beta_1$');

%repeat C
%estimate distribution of beta_hat

[f_beta0,x_beta0]=ksdensity(beta_hat4(:,2),'Kernel','epanechnikov','Bandwidth
    ',0.01);
figure;
plot(x_beta0,f_beta0);
xlabel('$\hat{\beta_0}$');
grid on;
title('Estimated p.d.f of $\beta_0$');

[f_beta1,x_beta1]=ksdensity(beta_hat4(:,1),'Kernel','epanechnikov','Bandwidth
    ',0.002);
figure;
plot(x_beta1,f_beta1);
xlabel('$\hat{\beta_1}$');
grid on;
title('Estimated p.d.f of $\beta_1$');
```

Exercise 3-Accounting for EIV When Forecasting Variance)

(A) Here is the summary table of MSE of different forecasting model.

Table 6: Summary Table of AR1,HAR1 and Non-change Model Forecasting

Stock	MSE_{AR1}	MSE_{HAR1}	MSE_{QAR1}	MSE_{QHAR1}	MSE_{NC}
PG	$1.2432 * 10^{-8}$	$1.1566 * 10^{-8}$	$1.0158 * 10^{-8}$	$1.0085 * 10^{-8}$	$1.6612 * 10^{-8}$
DIS	$1.1391 * 10^{-8}$	$9.8089 * 10^{-9}$	$1.0507 * 10^{-8}$	$1.0924 * 10^{-8}$	$1.2085 * 10^{-8}$

(B) According to the MSE table, the model with smallest MSE for PG and DIS is HARQ1 and HAR1 model, respectively. Even though there is no consistently better model for both stock.

Nice work! (+0.8)

(C) Here is the summary table of MSE of different forecasting model.

Table 7: Forecasting MSE of Different Models ($\times 10^{-8}$)

Stock	**AR1**	**HAR1**	**ARQ1**	**HARQ1**	**NC**
AAPL	3.2714	2.9762	2.7786	2.6902	3.9421
AXP	3.5744	1.5778	1.7709	1.4288	1.8856
BA	2.3167	2.1559	2.2445	2.1558	3.3353
BAC	12.7604	11.4649	12.1244	11.0409	13.0808
BLK	2.2313	1.7329	1.7581	1.6140	2.0247
C	12.1826	6.3003	6.8755	7.9979	5.7247
CAT	2.0648	1.8225	1.9354	1.8075	2.3188
CSCO	1.0665	0.9177	0.9491	0.8949	1.1621
CVX	1.3623	1.2868	1.3603	1.2812	1.4211
DIS	1.1391	0.9809	1.0507	1.0924	1.2085
GE	3.8628	3.3507	2.4765	2.4125	3.8388
GNTX	12.9547	11.1596	11.9238	10.9278	18.1811
GS	3.3944	2.1976	3.1675	3.5596	2.6432
HD	3.1479	2.6523	2.3320	2.2735	3.6952
IBM	0.6825	0.6393	0.5895	0.5903	0.7714
INTC	1.8806	1.6442	1.5698	1.5234	2.2521
JNJ	7.7365	8.0818	7.9765	8.1612	8.7594
JPM	3.4805	2.3850	2.5029	2.4623	3.0487
KO	0.3230	0.3134	0.2769	0.2841	0.3632
MCD	0.8678	0.7321	0.6505	0.6396	0.9998

Stock	AR1	HAR1	ARQ1	HARQ1	NC
MET	4.7331	3.4626	3.8582	3.3952	4.7734
MMC	1.0222	0.7447	0.7473	0.6641	0.8568
MMM	0.9572	0.7548	0.6677	0.6547	0.9877
MRK	1.3333	0.8740	0.8094	0.7869	1.0468
MS	15.0342	8.9090	47.5674	44.0217	8.9950
MSFT	1.9402	1.8601	1.6777	1.6693	2.3991
NKE	3.7374	3.6169	2.7173	2.7401	4.4002
PFE	1.2692	1.1040	1.2880	1.1690	1.5515
PG	1.2432	1.1566	1.0158	1.0085	1.6612
PNC	2.2143	1.1236	1.3913	1.2943	1.2884
SPY	0.5440	0.5224	0.4111	0.4313	0.5580
STT	26.1304	6.0592	4.7533	3.6078	2.3478
TSLA	3.6171	2.5752	3.4272	2.6248	3.5462
UNH	9.6769	8.7356	9.0138	8.6044	14.6258
UTX	0.7892	0.7204	0.7679	0.7382	0.9246
VZ	0.7412	0.6082	0.6622	0.6168	0.8915
WMT	1.6148	1.5385	1.5195	1.5042	2.5851
XOM	0.9464	0.8874	0.8611	0.8410	0.9395

From the MSE table we cannot find a model that is consistently better than other models for all stocks, however, we can find two models : HAR1 and Non-change, which are more stable than the AR1, ARQ1 and HARQ1 model for all stocks.

If QIV is much large for the data, using HARQ and ARQ model to reduce the estimate error in beta will be a bad idea since the large QIV will bring more biases for estimating.

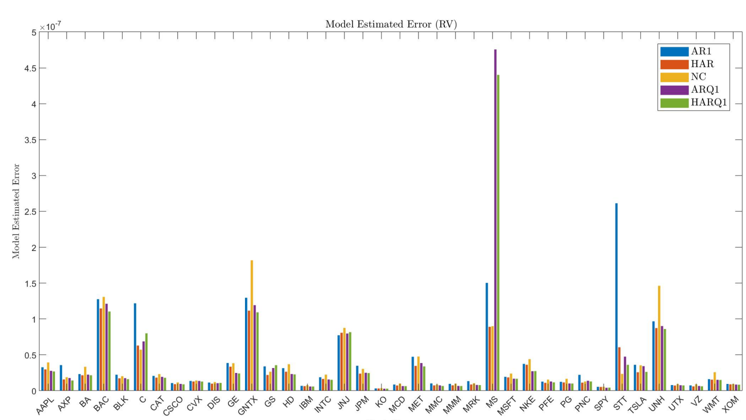


Figure 8: Bar Plot of RV Forecast MSE

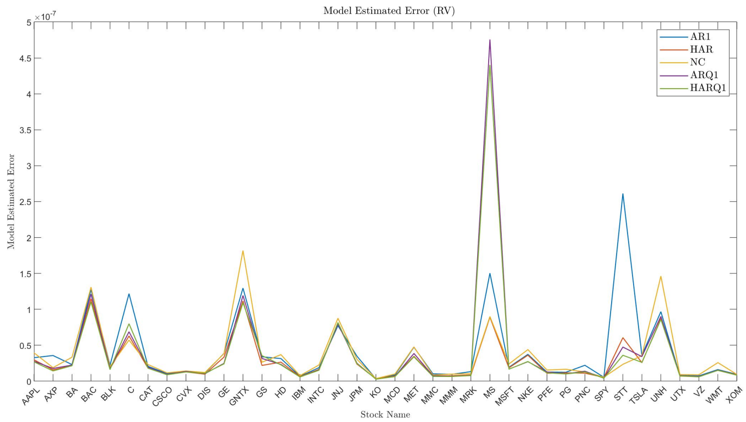


Figure 9: Figure of RV Forecast MSE

From the figures we can clearly find that for most of stocks, the difference of MSE of five models are small, which mean these five models work well in quasi-forecasting. However, for some specific stocks, like MS, QSR1 and QHAR1 work badly and they have about three time MSE as other's. For stock STT, AR1 model works badly and has two times of MSE as others.

In conclusion, HAR1 and Non-change model are the most stable model and work consistently better for all stocks.

(D) Here is the summary table of MSE of different forecasting model.

Table 8: Forecasting MSE of Different Models(TV)($\times 10^{-8}$)

Stock	AR1	HAR1	ARQ	HARQ1	NC
AAPL	3.2091	2.9623	2.7051	2.6453	3.8653
AXP	3.2987	1.4093	1.4958	1.2628	1.5763
BA	2.2521	2.0966	2.1955	2.1188	3.1431
BAC	12.5713	11.2341	12.0602	10.9838	13.0232
BLK	2.1816	1.7107	1.7118	1.5942	1.9908
C	12.6278	6.3383	6.5516	6.7494	5.6320
CAT	1.9681	1.7371	1.8305	1.7161	2.1880
CSCO	1.0136	0.9145	1.0763	0.9932	1.1534
CVX	1.3352	1.2672	1.3759	1.2812	1.3838
DIS	0.9758	0.8681	0.8626	0.8295	1.0034
GE	3.8282	3.3438	2.4326	2.3894	3.8257
GNTX	10.1214	8.6838	8.9482	8.3329	13.7327
GS	3.1909	2.1473	3.7833	3.8295	2.6105

Stock	AR1	HAR1	ARQ	HARQ1	NC
HD	3.1849	2.7063	2.3773	2.2965	3.6689
IBM	0.6501	0.6276	0.5924	0.5829	0.7544
INTC	1.4487	1.2924	1.2410	1.2159	1.7189
JNJ	0.2685	0.2613	0.2680	0.2580	0.2929
JPM	3.3475	2.3436	2.4872	2.4502	2.9863
KO	0.3173	0.3107	0.2783	0.2787	0.3566
MCD	0.8562	0.7186	0.6378	0.6286	0.9823
MET	4.1713	2.9094	3.2422	2.7849	3.7956
MMC	0.8604	0.7131	0.6189	0.6163	0.8430
MMM	0.8988	0.7535	0.7127	0.6753	0.9757
MRK	1.0049	0.7948	0.7572	0.7199	1.0017
MS	14.9350	8.3848	29.7105	29.0690	8.8612
MSFT	1.9118	1.8309	1.6274	1.6255	2.3419
NKE	3.6420	3.5443	2.5999	2.6164	4.2373
PFE	0.8766	0.7866	0.9020	0.8401	1.0316
PG	1.2324	1.1541	0.9917	0.9857	1.5692
PNC	2.1834	1.0978	1.3494	1.2580	1.2398
SPY	0.5394	0.5299	0.4119	0.4260	0.5483
STT	25.2062	5.9830	4.1699	3.3472	2.1732
TSLA	3.3020	2.3677	3.1266	2.4557	3.2622
UNH	9.4701	8.5657	8.7695	8.4156	14.3376
UTX	0.7187	0.6631	0.7325	0.6962	0.8210
VZ	0.7014	0.5765	0.6413	0.6027	0.8256
WMT	1.5999	1.5231	1.4871	1.4766	2.5414
XOM	0.9264	0.9125	0.8453	0.8185	0.9100

From the TV MSE table we can find, the range of TV forecast MSE is $(0.2929 \times 10^{-8}, 2.9711 \times 10^{-7})$, which is just a half of the range of RV MSE. Since TV just consider the diffusive return of stocks' log return, this will reduce the effects of jump returns and increase the accuracy of forecast.

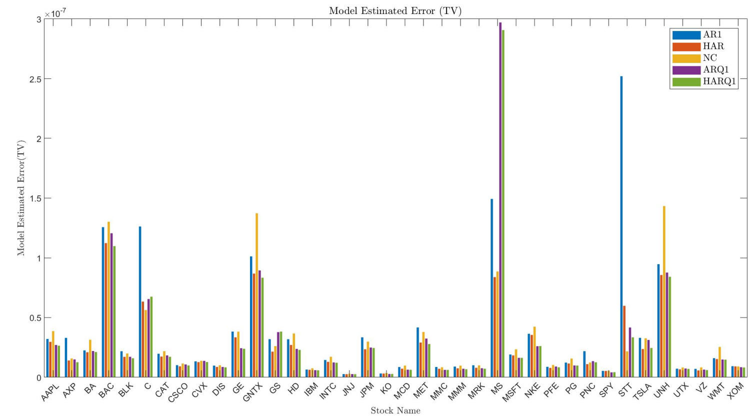


Figure 10: Bar Plot of TV Forecast MSE

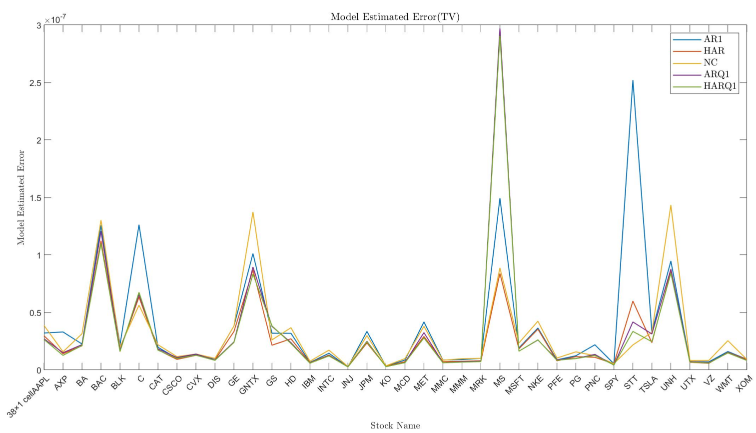


Figure 11: Figure of TV Forecast MSE

From the figures we can find that for most of stocks, these five models all work well and the difference between different models' MSE is very small. However, the same as the case in forecasting RV, there are stocks that some kind of models works badly. For example, HARQ and ARQ have large MSE for stock MS and AR1 model has the largest MSE for stock STT.

In conclusion, the MSE is smaller when forecasts RV and HAR1 and Non-change model are the most stable model and work consistently better for all stocks.

The **MATLAB**: Scripts of Q3 A-B

```
addpath('D:\ZM-Documents\MATLAB\data','functions','scripts');
[dates_PG,lp_PG]=load_stock('PG.csv','m');
N_PG=sum(floor(dates_PG(1,1))==floor(dates_PG(:,1)));% number of observations
     per day
T_PG=size(dates_PG,1)/N_PG;
[rdates_PG,lr_PG]=log_return([dates_PG lp_PG],N_PG,1);
days_PG=unique(floor(rdates_PG));

n=N_PG-1;
a=5;
[lr_c_PG,lr_d_PG]=c_d_log_returns(lr_PG,N_PG,a);

RV_PG=transpose(sum(lr_PG.^2));
QIV_PG=transpose(sum((n/3)*(lr_c_PG.^4)));

%3A
step=1;
X=RV_PG;
Y=RV_PG;
Qiv=QIV_PG;
J=1000-1;
fRV_adjAR1_PG=zeros(T_PG-J-1,1);%t increase
fRV_adjHAR1_PG=zeros(T_PG-J-1,1);
%fRV_NC_PG=zeros(T_PG-J);

for T=J+1:T_PG-1
    fRV_adjAR1_PG(T-J)=adjAR(X,Y,Qiv,J,T,step);
    fRV_adjHAR1_PG(T-J)=adjHAR(X,Y,Qiv,J,T,step);
end
ERR_adjAR1_PG=(mean((Y(J+2:T_PG)-fRV_adjAR1_PG).^2));
ERR_adjHAR1_PG=(mean((Y(J+2:T_PG)-fRV_adjHAR1_PG).^2));
ERR_CN_PG=(mean((Y(J+2:T_PG)-Y(J+1:T_PG-1)).^2));
```

Scripts of Q3 C

```
addpath('D:\ZM-Documents\MATLAB\data\Stocks5Min','functions','scripts');
Name=char('AAPL.csv','AXP.csv','BA.csv','BAC.csv','BLK.csv','C.csv','CAT.csv'
    ,...
    'CSCO.csv','CVX.csv','DIS.csv','GE.csv','GNTX.csv','GS.csv','HD.csv','IBM
       .csv','INTC.csv',...
    'JNJ.csv','JPM.csv','KO.csv','MCD.csv','MET.csv','MMC.csv','MMM.csv','MRK
       .csv','MS.csv','MSFT.csv','NKE.csv',...
    'PFE.csv','PG.csv','PNC.csv','SPY.csv','STT.csv','TSLA.csv','UNH.csv','
       UTX.csv','VZ.csv','WMT.csv','XOM.csv');
```

```
J=1000-1;
step=1;


ERR_AR=zeros(size(Name,1),1);
ERR_HAR=zeros(size(Name,1),1);
ERR_adjAR=zeros(size(Name,1),1);
ERR_adjHAR=zeros(size(Name,1),1);
ERR_NC=zeros(size(Name,1),1);

for i=1:size(Name,1)
[ERR_AR(i),ERR_HAR(i),ERR_adjAR(i),ERR_adjHAR(i),ERR_NC(i)] = MSE(strtrim(
    Name(i,:)),J,step);
end

ERR=[ERR_AR,ERR_HAR,ERR_NC,ERR_adjAR,ERR_adjHAR];
figure;
bar(ERR);
xlabel('Stock Name');
ylabel('Model Estimated Error');
title('Model Estimated Error');
legend('AR1','HAR','NC','ARQ1','HARQ1');

figure;
plot(1:38,ERR,'linewidth',1);
xlabel('Stock Name');
xlim([1,38]);
ylabel('Model Estimated Error');
title('Model Estimated Error');
legend('AR1','HAR','NC','ARQ1','HARQ1');
```

Scripts of Q3 D

```
addpath('D:\ZM-Documents\MATLAB\data\Stocks5Min','functions','scripts');
Name=char('AAPL.csv','AXP.csv','BA.csv','BAC.csv','BLK.csv','C.csv','CAT.csv'
    ,...
    'CSCO.csv','CVX.csv','DIS.csv','GE.csv','GNTX.csv','GS.csv','HD.csv','IBM
        .csv','INTC.csv',...
    'JNJ.csv','JPM.csv','KO.csv','MCD.csv','MET.csv','MMC.csv','MMM.csv','MRK
        .csv','MS.csv','MSFT.csv','NKE.csv',...
    'PFE.csv','PG.csv','PNC.csv','SPY.csv','STT.csv','TSLA.csv','UNH.csv','
        UTX.csv','VZ.csv','WMT.csv','XOM.csv');
J=1000-1;
step=1;

```

```

ERR_AR_T=zeros(size(Name,1),1);
ERR_HAR_T=zeros(size(Name,1),1);
ERR_adjAR_T=zeros(size(Name,1),1);
ERR_adjHAR_T=zeros(size(Name,1),1);
ERR_NC_T=zeros(size(Name,1),1);

for i=1:size(Name,1)
[ERR_AR_T(i),ERR_HAR_T(i),ERR_adjAR_T(i),ERR_adjHAR_T(i),ERR_NC_T(i)] =
    MSE_TV(strtrim(Name(i,:)),J,step);
end

ERR=[ERR_AR_T,ERR_HAR_T,ERR_NC_T,ERR_adjAR_T,ERR_adjHAR_T];
figure;
bar(ERR);
xlabel('Stock Name');
ylabel('Model Estimated Error(TV)');
title('Model Estimated Error');
legend('AR1','HAR','NC','ARQ1','HARQ1');

figure;
plot(ERR,'linewidth',1);
xlabel('Stock Name');
xlim([1,38]);
ylabel('Model Estimated Error');
title('Model Estimated Error(TV)');
legend('AR1','HAR','NC','ARQ1','HARQ1');
```