90.5/100

Ziming Huang

Midterm

Econ 672
October 3, 2019

Question 0

We will use these stocks to solve Question 1 to 3.

Stock	Ticker
1	PG
2	DIS
3	SPY

Question 1

(A) (i) In our data set, September 16, 2008 is the 430_{th} observation day, so $t(September 16, 2008) = 430$.

(ii) It is not possible to use today's data to estimate today's IV, since it is the morning now, we do not have enough price data to do this estimation. However, we can use the historical data to estimate today's IV by doing forecasting.

(B) The follow is the estimated value of IV in September 15, 2008.

Stock	Estimated IV of September 15, 2008(using TV)
PG	2.2430×10^{-4}
DIS	4.9671×10^{-4}

(i) Since there are jumps existing in stock price, to better estimate the value of IV, we here choose truncated variance as a estimator for IV. This value measures the intraday fluctuation in stock's price: the bigger value of estimated IV, the more volatility we will expected for stock's price. From this table, the DIS' IV_t is larger than PG's, which means the price(or return) volatility of DIS is bigger than PG's, that is, DIS' stock is more risky.

(ii) The reason we use the historical (yesterday) data to do forecasting for today's IV is that: the history data reflects the information of a company, this information includes its financial performance, financial position and operation strategy. Besides, the history data also contain the relationship between the company and the market environment, such as market interest, CPI index and other micro-structure shocks. This information is really important for us to understand and evaluate a company. Since a company's structure and relationship between market will not change in a short time, it is reasonable for us to believe that the recent history data can reflect company's status well so it can be use to do future forecasting. (we still need to think carefully whether this assumption holds or not: the future data shares the same distribution of historical data. Only under this assumption can we regard historical data as a good estimator for future statistics and these forecasts will be efficient).

(C) To calculate the loss probability, we follow these steps:

(i) Estimate the parameter IV_t in the r_t^c 's asymptotic distribution $\mathcal{N}(0, IV_t)$ by truncated variance TV_t;

(ii) Standardize r_t^c and compute the loss probability by using the CDF of standard normal distribution:

$$\frac{r_t^c}{\sqrt{IV_t}} = Z_t \sim \mathcal{N}(0,1)$$

$$\mathbb{P}(r_t^c <= -q) = \mathbb{P}(\frac{r_t^c}{\sqrt{IV_t}} <= \frac{-q}{\sqrt{IV_t}}) = \mathbb{P}(Z_t <= \frac{-q}{\sqrt{IV_t}}) = 1 - \Phi(\frac{q}{\sqrt{IV_t}})$$

By replacing q=2% (or 4%) and $\hat{IV_t} = TV_{yesterday}$, we can compute the loss probability.

To calculate the $CI(IV_t)$ and $CI(\mathbb{P})$, we follow these steps:

(i) Compute $CI(\Delta_n^{-0.5}(TV_t - IV_t))$ from its asymptotic distribution:

$$\Delta_n^{-0.5}(TV_t - IV_t) \xrightarrow{d} \mathcal{N}(0, 2QIV_t)$$

$$\Rightarrow \quad CI(\Delta_n^{-0.5}(TV_t - IV_t)) = [-c\sqrt{2QIV_t}, \;\; c\sqrt{2QIV_t}\;]$$

(ii) Deduce the $CI(IV_t)$'s expression from $CI(\Delta_n^{-0.5}(TV_t - IV_t))$:

$$CI(IV_t, c) = [-c\sqrt{2\Delta_n QIV_t} + TV_t, \;\; c\sqrt{2\Delta_n QIV_t} + TV_t]$$

(iii) Estimate parameter QIV_t and compute the CI of IV_t:

$$Q\hat{I}V_t = \frac{1}{3\Delta_n}\sum_{i=1}^{n}(r_{t,i}^c)^4$$

$$CI(IV_t) = [-c\sqrt{2\Delta_n Q\hat{I}V_t} + TV_t, \;\; c\sqrt{2\Delta_n Q\hat{I}V_t} + TV_t]$$

where c is the critical value and equals $\Phi(1 - \frac{0.05}{2}) = 1.96$ in this case (95% significant level).

(iv) Use $CI(IV_t)$'s upper bound and lower bound as new variance in r_t^c 's asymptotic distribution to calculate $CI(\mathbb{P})$:

$$CI^{up}(\mathbb{P}) = \mathbb{P}(r_t^c <= -q) = \mathbb{P}(Z_t <= \frac{-q}{\sqrt{IV_t^{upper}}}) = 1 - \Phi(\frac{q}{\sqrt{IV_t^{upper}}})$$

$$CI^{low}(\mathbb{P}) = \mathbb{P}(r_t^c <= -q) = \mathbb{P}(Z_t <= \frac{-q}{\sqrt{IV_t^{lower}}}) = 1 - \Phi(\frac{q}{\sqrt{IV_t^{lower}}})$$

By replacing q=2% (or 4%), we can compute the $CI(\mathbb{P})$.

		Estimate of $\mathbb{P}(r_t^c <= -q)$		
Stock	q	Estimate	Lower bound	Upper bound
PG	2%	0.0909	0.0404	0.1308
PG	4%	0.0038	2.3953×10^{-4}	0.0124
DIS	2%	0.1848	0.1392	0.2169
DIS	4%	0.0363	0.0151	0.0588

According to the table, we can find that the width of confidence interval is in the range 0.01-0.09, which means the biggest difference between real probability of losing certain amount of money and the estimated probability will be 9 percentage. This outcome isn't very accuracy. However, given that confidence intervals' width of losing 4% money fall in the range 0.1-0.4, this estimated confidence interval may be accuracy in this case.

(D) To compute the VaR and its confidence interval, we follow these steps:

(i) Transfer the asymptotic distribution of r_t^c to standard normal distribution:

$$\frac{r_t^c}{\sqrt{IV_t}} = Z_t \sim \mathcal{N}(0,1)$$

(ii) Compute the loss probability by the CDF of standard normal distribution(here r_t^c % is the percentage log return):

$$\mathbb{P}(\frac{r_t^c\ \%}{100} \times V <= Q) = \mathbb{P}(r_t^c\ \% <= \frac{100Q}{V}) = \Phi(\frac{100Q}{V\sqrt{IV_t}}) = p$$

(iii) Compute the value of Q by taking inverse to the CDF of standard normal distribution:

$$Q = \Phi^{-1}(p)\frac{V\sqrt{IV_t}}{100}$$

(iv) (*Delta Method*) Since we have already known the asymptotic distribution of $\Delta_n^{-0.5}(TV_t - IV_t)$, we can apply Delta Method to deduce the asymptotic distribution of Q and then use it to compute $CI(Q)$: wrong way of deriving delta method, the derivative of normal pdf should be included (-2)

$$\Delta_n^{-0.5}(\sqrt{TV_t} - \sqrt{IV_t}) \xrightarrow{d} \mathcal{N}(0, 2QIV_t)$$

$$Delta\ Method \Rightarrow \quad \Delta_n^{-0.5}\Phi^{-1}(p)\frac{V}{100}(\sqrt{TV_t} - \sqrt{IV_t}) \xrightarrow{d} \mathcal{N}(0, \frac{QIV_t}{2IV_t})$$

so the confidence interval of Q is:

$$CI^{up}(Q, 1-\alpha) = \Phi^{-1}(p)\frac{V}{100}(\sqrt{TV_t}) + c * \Phi^{-1}(p)\frac{V}{100}\sqrt{\frac{\Delta_n\hat{Q}IV_t}{2TV_t}}$$
$$CI^{low}(Q, 1-\alpha) = \Phi^{-1}(p)\frac{V}{100}(\sqrt{TV_t}) - c * \Phi^{-1}(p)\frac{V}{100}\sqrt{\frac{\Delta_n\hat{Q}IV_t}{2TV_t}}$$

VaR Estimate and 95% Confidence interval				
Stock	p	Estimate	Lower bound	Upper bound
PG	1%	-6.9681	-5.5126	-8.4236
PG	5%	-4.9268	-3.8977	-5.9560
DIS	1%	-10.3694	-8.7269	-12.0119
DIS	5%	-7.3318	-6.1704	-8.4931

take care the positions of number(-3)

According to the table, we can find that the width of confidence interval is in the range 2.0 - 3.3. To cancel the effect of scale, it is much better to use percentage width to evaluate the accuracy of confidence interval. In thus case, the percentage width of confidence interval is in the range of 1.00% - 1.65%, so the level of accuracy is sufficient and my boss may find them helpful.

(Simplified)

We can also use $CI^{up}(IV_t)$ and $CI^{low}(IV_t)$ as new variance to compute Q. These two values of Q will be our $CI^{up}(Q)$ and $CI^{low}(Q)$:

$$CI^{up}(Q) = \Phi^{-1}(p)\frac{V\sqrt{IV_t^{upper}}}{100}$$

$$CI^{low}(Q) = \Phi^{-1}(p)\frac{V\sqrt{IV_t^{lower}}}{100}$$

VaR Estimate and 95% Confidence interval				
Stock	p	Estimate	Lower bound	Upper bound
PG	1%	-6.9681	-5.3292	-8.2891
PG	5%	-4.9268	-3.7681	-5.8608
DIS	1%	-10.3694	-8.5837	-11.8899
DIS	5%	-7.3318	-6.0692	-8.4068

According to the table, the estimated confidence interval here is very close to what we get by Delta method. The width of confidence interval here is a little bigger:from 2.1 to 3.3, the range percentage to initial investment V=200 is 1.05%-1.65%, this indicates that using the simplified method to estimate confidence interval can also give us a precise level of accuracy which efficient in application of prediction.

The **MATLAB** code:

```
%——Q1

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)));%#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));
alpha=5;
[lr_c_PG,lr_d_PG]=c_d_log_returns(lr_PG,N_PG-1,alpha);

%part A
%pick yesterday September 15,2008
d=max(find(floor(rdates_PG)==datenum('20080915','yyyymmdd')))/77;%find
    yestoday's data

%part B
%yesterday's TV to estimate IV
lr_c_PG1=lr_c_PG(:,d);% get yestoday's market data
TV_PG1=truncated_var_day(lr_c_PG1); % calculate yestoday's TV
QIV_PG1=sum(((N_PG-1)/3)*(lr_c_PG1.^4));

%part C
%loss probability
pd = makedist('Normal','mu',0,'sigma',1);
p2_PG = 1-cdf(pd,0.02/sqrt(TV_PG1));
p4_PG = 1-cdf(pd,0.04/sqrt(TV_PG1));

%95% CI of loss probability

%CI of IV
c=-norminv(0.025);
TVl_PG1=TV_PG1-c*sqrt(2*QIV_PG1/N_PG);
TVu_PG1=TV_PG1+c*sqrt(2*QIV_PG1/N_PG);
%CI of probability by using IV's CI
%q=0.02
p2u_PG = 1-cdf(pd,0.02/sqrt(TVu_PG1));
p2l_PG = 1-cdf(pd,0.02/sqrt(TVl_PG1));
%q=0.04
p4u_PG = 1-cdf(pd,0.04/sqrt(TVu_PG1));
p4l_PG = 1-cdf(pd,0.04/sqrt(TVl_PG1));

%part D
%VaR
V=200;% million
Q1_PG=norminv(0.01)*V*sqrt(TV_PG1); %lr here is not percentage
Q5_PG=norminv(0.05)*V*sqrt(TV_PG1);
```

```
%CI of VaR

%delta method
g1TV_PG=norminv(0.01)*V*sqrt(TV_PG1);
g5TV_PG=norminv(0.05)*V*sqrt(TV_PG1);

c=-norminv(0.025);
gQ1u_PG=g1TV_PG+c*norminv(0.01)*V*sqrt(QIV_PG1/(2*(N_PG-1)*TV_PG1));
gQ1l_PG=g1TV_PG-c*norminv(0.01)*V*sqrt(QIV_PG1/(2*(N_PG-1)*TV_PG1));
gQ5u_PG=g5TV_PG+c*norminv(0.05)*V*sqrt(QIV_PG1/(2*(N_PG-1)*TV_PG1));
gQ5l_PG=g5TV_PG-c*norminv(0.05)*V*sqrt(QIV_PG1/(2*(N_PG-1)*TV_PG1));

%simplified method
Q1u_PG=norminv(0.01)*V*sqrt(TVu_PG1);
Q1l_PG=norminv(0.01)*V*sqrt(TVl_PG1);
Q5u_PG=norminv(0.05)*V*sqrt(TVu_PG1);
Q5l_PG=norminv(0.05)*V*sqrt(TVl_PG1);
```

Exercise 2

In this question, the long stock A is PG and the short stock B is DIS.

(A) The follow is the statistic summary table. not so close (-0.5)

Summary Statistics of Long-Short Continuous returns				
Average	Min	5% Percentile	95% Percentile	Max
1.5566×10^{-6}	-0.0416	-0.0021	0.0022	0.0357

(B) The follow is the statistic summary table of realized beta of long-short portfolio.

Summary Statistics of the Realized Beta				
Average	Min	5% Percentile	95% Percentile	Max
-0.3899	-2.2904	-0.9650	0.2081	1.3388

(C) Here is the figure of realized beta and CI of of long-short portfolio from 2007 to 2017.

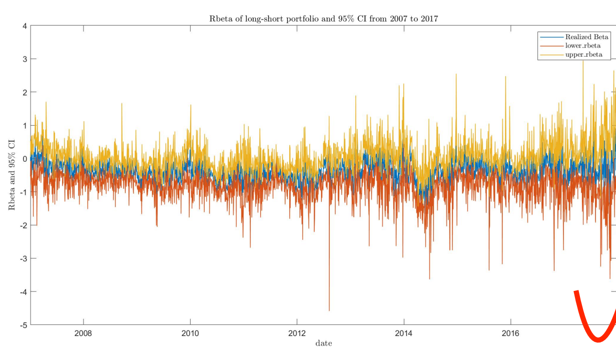

Figure 1: Rβ and estimated CI of long-short portfolio from 2007 to 2017

From this figure we can see most of the value of long-short portfolio concentrate on zero, which means this portfolio may show the character of market neutral. However, there exists some periods that the absolute values of realized beta are large: around 2010, 2013, 2014 and 2017, which are consistent with financial crisis. This discover may be an evidence to support the hypnosis that stocks correlation to market may increase during crisis.

(D) The follow is the figure of realized beta and estimated CI of long-short portfolio in October 2008:

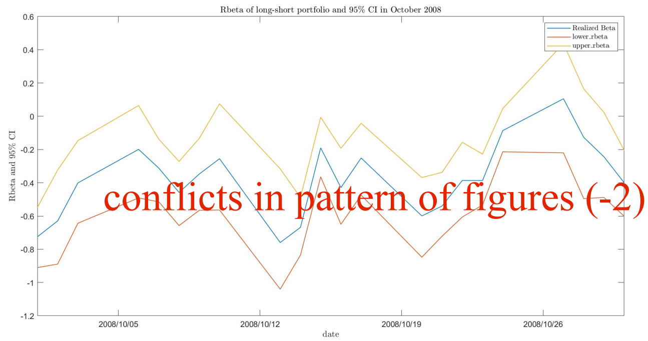

Figure 2: Rβ and estimated CI of long-short portfolio in October 2008

From this figure we can find the confidence interval estimated by bootstrap is good for this specific period: all the observation data fall into the region bounded by lower confidence interval and upper confidence interval. However, the range of confidence interval width in this period is 0.15-0.6, which means the confidence interval here is not precise enough for most application.

As for the value of confidence interval and realized beta, we can see that about 80% of observations and the estimated confidence intervals are below 0, which means this new portfolio may not be market neutral and it shows negative correlation to market, though this correlation is not weak (most of the absolute value of realized beta is larger than 0.3).

(E) To better estimate whether this new portfolio is market neutral or not, we calculate the number of confidence intervals that contains 0.

Summary Statistics of confidence interval		
Types	Number	Percentage(%)
Do not reject $H_0 : R\beta = 0$ ($0 \in CI$)	1352	48.8263
Reject H_0 and in favor of $R\beta > 0$ (CI_{low}>0)	16	0.5778
Reject H_0 and in favor of $R\beta > 0$ (CI_{up}<0)	1401	50.5959

According to the statistical summary table of confidence interval, we can see that the number of confidence intervals contain 0 is 1352 and takes up 48.8263% of total sample. This result is not enough for us to reach a conclusion that this new portfolio is market neutral even though the figure shows some clues.

not so close (-1)

The **MATLAB** code:

```
%Question 2

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)));%#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));

[dates_DIS,lp_DIS]=load_stock('DIS.csv','m');
N_DIS=sum(floor(dates_DIS(1,1))==floor(dates_DIS(:,1)));%#observations per
    day
T_DIS=size(dates_DIS,1)/N_DIS;
[rdates_DIS,lr_DIS]=log_return([dates_DIS lp_DIS],N_DIS,1);
days_DIS=unique(floor(rdates_DIS));

[dates_SPY,lp_SPY]=load_stock('SPY.csv','m');
N_SPY=sum(floor(dates_SPY(1,1))==floor(dates_SPY(:,1)));%#observations per
    day
T_SPY=size(dates_SPY,1)/N_SPY;
[rdates_SPY,lr_SPY]=log_return([dates_SPY lp_SPY],N_SPY,1);
days_SPY=unique(floor(rdates_SPY));

alpha=5;
[lr_c_PG,lr_d_PG]=c_d_log_returns(lr_PG,N_PG-1,alpha);
[lr_c_DIS,lr_d_DIS]=c_d_log_returns(lr_DIS,N_DIS-1,alpha);
[lr_c_SPY,lr_d_SPY]=c_d_log_returns(lr_SPY,N_SPY-1,alpha);


% long asset is PG and short asset is DIS

%part A
%summary long-short continuous returns
lr_p=lr_PG-lr_DIS;%return of long-short portfolio
[lr_c_p,lr_d_p]=c_d_log_returns(lr_p,N_PG-1,alpha);
[m_1,mi_1,ma_1,q1_1,q2_1]=summary(lr_c_p(:));

%part B
% realized beta
rbeta=realized_beta(lr_c_p,lr_c_SPY);
[m_2,mi_2,ma_2,q1_2,q2_2]=summary(rbeta);

%part C
%compute CI of rbeta
```

```
kn=11;
n=N_PG-1;
r=1000;

sbeta=zeros(r,T_PG);
parfor i=1:r
    [newsample1,newsample2]=bsample_2(lr_c_p,lr_c_SPY,kn,n,T_PG);
    sbeta(i,:)=sum(newsample1.*newsample2)./sum(newsample2.^2);
end
%CI of rbeta
CI_low_rbp = quantile(sbeta, 0.025);
CI_up_rbp = quantile(sbeta, 0.975);

%plot long-short rbeta and its CI
figure;
plot(days_PG,rbeta);
hold on;
plot(days_PG,CI_low_rbp);
hold on;
plot(days_PG,CI_up_rbp);
legend('Realized Beta','lower\_rbeta','upper\_rbeta');
xlabel('date');
ylabel('Rbeta and 95\% CI');
title('Rbeta of long-short portfolio and 95\% CI from 2007 to 2017');
datetick('x','keeplimits');
xlim([min(days_PG),max(days_PG)]);
hold off;

%part D
num1=datenum('20081001','yyyymmdd');
num2=datenum('20081031','yyyymmdd');
a=sum(rdates_PG(:)<num1)/77;
b=sum(rdates_PG(:)<=num2)/77;
CI_low_rbp_1=CI_low_rbp(a+1:b);
CI_up_rbp_1=CI_up_rbp(a+1:b);
rbeta_1=rbeta(a+1:b);

%plot rbeta and CI in October 2008
figure;
plot(days_PG(a+1:b),rbeta_1);
hold on;
plot(days_PG(a+1:b),CI_low_rbp_1);
hold on;
plot(days_PG(a+1:b),CI_up_rbp_1);
```

```
legend('Realized Beta','lower\_rbeta','upper\_rbeta');
xlabel('date');
ylabel('Rbeta and 95\% CI');
title('Rbeta of long-short portfolio and 95\% CI in October 2008');
datetick('x',26,'keeplimits');
xlim([min(days_PG(a+1:b)),max(days_PG(a+1:b))]);
hold off;


%part E
%formal evaluation
%construct confidence interval for 0
%evaluate market neutral
n1=sum(CI_low_rbp <=0 & CI_up_rbp >=0);% CI contain 0
n2=sum( CI_up_rbp < 0);%CI lower 0
n3=sum( CI_low_rbp > 0);%CI upper 0
```

Exercise 3

(A) The follow is the statistic summary table of continuous return of PG,DIS and SPY.

Summary Statistics for Continuous Returns

Stock	Average	Min	5% Percentile	95% Percentile	Max
PG	5.4237×10^{-6}	-0.0254	-0.0015	0.0016	0.0420
DIS	4.3967×10^{-6}	-0.0323	-0.0022	0.0022	0.0474
SPY	1.7039×10^{-6}	-0.0293	-0.0015	0.0015	0.0376

(B) The number of detected jumps in the market is 116.

(C) The follow is the statistic summary table of realized beta of PG and DIS:

Summary Statistics of Realized Beta

Stock	Average	Min	5% Percentile	95% Percentile	Max
PG	0.5403	-0.4167	0.1472	0.9709	2.0421
DIS	0.9312	-1.0811	0.4731	1.3581	2.8132

(D) Here are the figures of realized beta of PG and DIS from 2007 to 2017.

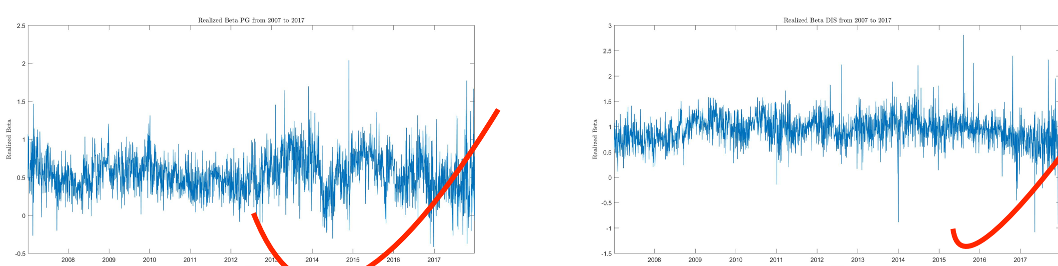

Figure 3: Realized beta and its estimated confidence interval of PG and DIS

From the figures, PG's realized beta mostly falls below the line $R\beta = 1$, which means that PG may have a lower beta than the market for most in trading day. The situation is different in the case of DIS. As we can see from the figure, most of DIS' realized beta falls above or around the line $R\beta = 1$, this indicates that the beta of DIS may have a higher or equal value to the market's in most of trading day.

We can do some statistical work to support our viewpoint. Here is the summary table of estimated realized beta for PG and DIS.

Summary Statistics of confidence interval			
Stock	Types	Number	Percentage(%)
PG	$CI^l > 1$	2	0.0722
	$1 \in CI$	665	24.0159
	$CI_u < 1$	2102	75.9119
DIS	$CI^l > 1$	204	7.3673
	$0 \in CI$	2100	75.8397
	$CI_u < 1$	465	16.7930

For PG stock, the number of confidence interval below 1 takes up 75.91%, which indicates PG stock will probably have a lower beta than the market for most of its trading day. A lower beta value means that the stock is less volatility than the market and has a lower risk and returns. For DIS stock, the number of confidence interval contain 1 takes up 75.84%, which indicates DIS stock will probably have a equal beta with the market for most of its trading day. The beta value equals to market beta means that the stock has the same volatility with the market.

(E) The follow is the statistic summary table of realized beta of PG and DIS:

Summary Statistics of Residual Correlation				
Average	Min	5% Percentile	95% Percentile	Max
0.0186	-0.4993	-0.2443	0.2881	0.5563

According to this table, the correlation of PG and DIS' residual PG and DIS is really small, which means that the stock returns for both stocks are mostly explained by the market factor and they are not very related to each other.

(F) Here is the figure of correlation and its 95% estimated confidence interval.

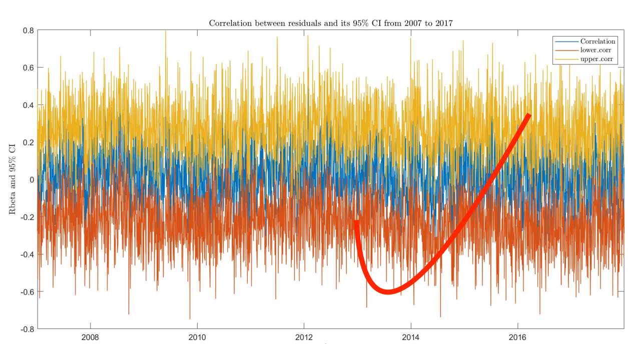

Figure 4: Correlation and its 95% CI from 2007 to 2017

From the figure we can see the correlations between two residuals are concentrated on zero and most of the confidence intervals have covered value zero, which means the linear relation between this two residuals is weak. Since these residuals are produced by removing the common movement with market from their log-returns, the weak linear relationship between residuals may indicate that these two stocks do not share other common movements except the market co-movement.

(G) Here is the figure of correlation and its 95% estimated confidence interval for October 2008.

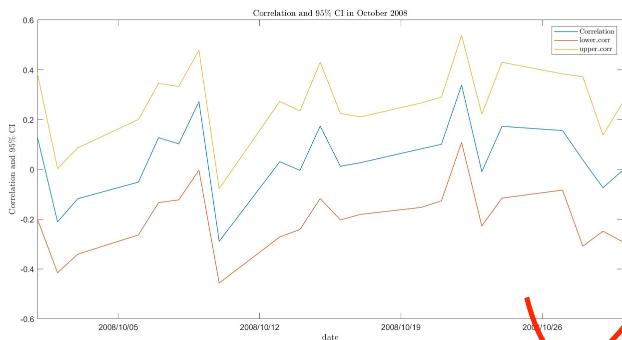

Figure 5: Correlation and its 95% CI for October 2008

According to the figure, the range of correlation between residuals is (-0.3, 0.3) and most of them are moving within (-0.1, 0.1). This indicates that the correlation is weak between these to variables. As for the confidence interval, the width is in the range of (0.1, 0.5), the max and min value of confidence interval is 0.5 and -0.4 respectively. From these data we can see that the possible correlation between the residuals will not be very strong, it is reasonable to reach the conclusion that the two stocks do not share other common movements except for the market movement.

(H) To better understand the correlation between residuals, we calculate the number of correlation confidence intervals that contains 0.

slight difference (-1)

Summary Statistics of confidence interval

Types	Number	Percentage(%)	"p"
Do not reject $H_0 : \rho_{et}$=0:	2501	90.3214	p_0=9.9246
Reject H_0 and in favor of $\rho_{et} > 0$	168	6.0672	p_+=0.6667
Reject H_0 and in favor of $\rho_{et} < 0$	100	3.6114	p_-=0.3968

According to the statistical summary table of confidence interval, we can see that the number of confidence interval that contain 0 is 2501 and takes up 90.3214% of total sample. And the p_0 value here represent the possible number of years in our data that the two stock do not have other co-movements except the market movements. As we can see that, from total observe time length (2007 to 2017), the number of year can not reject $\rho_{et} = 0$ is 9.92, that is for most of our observation, the correlation of residuals is near to zero. All these results support us to reach a conclusion that these two stocks do not have other common movements.

The **MATLAB** code:

```
%Question 3

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));

[dates_DIS,lp_DIS]=load_stock('DIS.csv','m');
N_DIS=sum(floor(dates_DIS(1,1))==floor(dates_DIS(:,1)));% number of
    observations per day
T_DIS=size(dates_DIS,1)/N_DIS;
[rdates_DIS,lr_DIS]=log_return([dates_DIS lp_DIS],N_DIS,1);
days_DIS=unique(floor(rdates_DIS));

[dates_SPY,lp_SPY]=load_stock('SPY.csv','m');
N_SPY=sum(floor(dates_SPY(1,1))==floor(dates_SPY(:,1)));% number of
    observations per day
T_SPY=size(dates_SPY,1)/N_SPY;
[rdates_SPY,lr_SPY]=log_return([dates_SPY lp_SPY],N_SPY,1);
days_SPY=unique(floor(rdates_SPY));

alpha=5;
[lr_c_PG,lr_d_PG]=c_d_log_returns(lr_PG,N_PG-1,alpha);
[lr_c_DIS,lr_d_DIS]=c_d_log_returns(lr_DIS,N_DIS-1,alpha);
[lr_c_SPY,lr_d_SPY]=c_d_log_returns(lr_SPY,N_SPY-1,alpha);



%part A
%summary of continuous returns
[pgm,pgmi,pgma,pgq1,pgq2]=summary(lr_c_PG(:));
```

```
[dism,dismi,disma,disq1,disq2]=summary(lr_c_DIS(:));
[spym,spymi,spyma,spyq1,spyq2]=summary(lr_c_SPY(:));

%part B
%number of jumps in the marhet
numj_spy=sum(sum(lr_d_SPY~=0));%pick jumps

%part C
%realized beta
rbeta_PG=realized_beta(lr_c_PG,lr_c_SPY);
rbeta_DIS=realized_beta(lr_c_DIS,lr_c_SPY);
%summary of rbeta
[pgrbm,pgrbmi,pgrbma,pgrbq1,pgrbq2]=summary(rbeta_PG);
[disrbm,disrbmi,disrbma,disrbq1,disrbq2]=summary(rbeta_DIS);

%part D
%plot realized beta
%PG
figure;
plot(days_PG, rbeta_PG);
xlabel('year');
ylabel('Realized Beta');
xlim([min(days_PG),max(days_PG)]);
title('Realized Beta PG from 2007 to 2017');
datetick('x','keeplimits');
%DIS
figure;
plot(days_DIS, rbeta_DIS);
xlabel('year');
ylabel('Realized Beta');
xlim([min(days_PG),max(days_PG)]);
title('Realized Beta DIS from 2007 to 2017');
datetick('x','keeplimits');

%evaluate CI
% CI of rbeta
kn=11;
n=N_PG-1;
r=1000;

%PG
sbeta_PG=zeros(r,T_PG);
parfor i=1:r
    [newsample1,newsample2]=bsample_2(lr_c_PG,lr_c_SPY,kn,n,T_PG);
```

```
    sbeta_PG(i,:)=sum(newsample1.*newsample2)./sum(newsample2.^2);
end
%CI of PG's rbeta
CI_low_rbPG = quantile(sbeta_PG, 0.025);
CI_up_rbPG = quantile(sbeta_PG, 0.975);

n1_PG=sum(CI_low_rbPG <=1 & CI_up_rbPG >=1);
n2_PG=sum( CI_up_rbPG < 1);
n3_PG=sum( CI_low_rbPG > 1);


%DIS
sbeta_DIS=zeros(r,T_DIS);
parfor i=1:r
    [newsample1,newsample2]=bsample_2(lr_c_DIS,lr_c_SPY,kn,n,T_DIS);
    sbeta_DIS(i,:)=sum(newsample1.*newsample2)./sum(newsample2.^2);
end
%CI of DIS's rbeta
CI_low_rbDIS = quantile(sbeta_DIS, 0.025);
CI_up_rbDIS = quantile(sbeta_DIS, 0.975);

n1_DIS=sum(CI_low_rbDIS <=1 & CI_up_rbDIS >=1);
n2_DIS=sum( CI_up_rbDIS < 1);
n3_DIS=sum( CI_low_rbDIS > 1);

%part E
%residual of non-market part
rbeta_PG1=repmat(rbeta_PG,N_PG-1,1);
rbeta_DIS1=repmat(rbeta_DIS,N_DIS-1,1);
e_PG=lr_c_PG-rbeta_PG1.*lr_c_SPY;
e_DIS=lr_c_DIS-rbeta_DIS1.*lr_c_SPY;
rho=diag(corr(e_PG,e_DIS)); % corrlation of same day's e_PG and e_DIS
[rm,rmi,rma,rq1,rq2]=summary(rho);

%part F
%bootstrapt to estimate CI(corr)
kn=11;
n=N_PG-1;
r=1000;

scorr=zeros(r,T_PG);
parfor i=1:r
    [PG,DIS,SPY]=bsample_3(lr_c_PG,lr_c_DIS,lr_c_SPY,kn,n,T_PG);
    e1=PG-repmat(sum(PG.*SPY)./sum(SPY.^2),N_PG-1,1).*SPY;
```

```
    e2=DIS−repmat(sum(DIS.*SPY)./sum(SPY.^2),N_DIS−1,1).*SPY;
    rho1=diag(corr(e1,e2));
    scorr(i,:)=rho1;
end

%CI of corr
CI_low_corr = quantile(scorr, 0.025);
CI_up_corr = quantile(scorr, 0.975);

%plot corr and its CI
figure;
plot(days_PG,rho);
hold on;
plot(days_PG,CI_low_corr);
hold on;
plot(days_PG,CI_up_corr);
legend('Correlation','lower\_corr','upper\_corr');
xlabel('date');
ylabel('Rbeta and 95\% CI');
title('Correlation between residuals and its 95\% CI from 2007 to 2017');
datetick('x','keeplimits');
xlim([min(days_PG),max(days_PG)]);
hold off;

%part G
%zoom in plot of corr and its CI
num1=datenum('20081001','yyyymmdd');
num2=datenum('20081031','yyyymmdd');
a=sum(rdates_PG(:)<num1)/77;
b=sum(rdates_PG(:)<=num2)/77;
CI_low_corr_1=CI_low_corr(a+1:b);
CI_up_corr_1=CI_up_corr(a+1:b);
rho_1=rho(a+1:b);

figure;
plot(days_PG(a+1:b),rho_1);
hold on;
plot(days_PG(a+1:b),CI_low_corr_1);
hold on;
plot(days_PG(a+1:b),CI_up_corr_1);
legend('Correlation','lower\_corr','upper\_corr');
xlabel('date');
ylabel('Correlation and 95\% CI');
title('Correlation and 95\% CI in October 2008');
```

```
datetick('x',26,'keeplimits');
xlim([min(days_PG(a+1:b)),max(days_PG(a+1:b))]);
hold off;

%part H
% construct confidence interval for 0
%evaluate common variance
n0=sum(CI_low_corr <=0 & CI_up_corr >=0);
nm=sum( CI_up_corr < 0);
np=sum( CI_low_corr > 0);

p0=n0/252; % do not reject H0
pm=nm/252; % reject and in favor of corr<0
pp=np/252; % reject and in favor of corr>0
```