Ziming Huang

Project 7

Econ 672
November 1, 2019

Exercise

(A) The coarse sample function is as follow: **AR Model**

```
function [log_returns] = log_return_new(log_price,kn,s,e)
%the input data log_price is a N*T matrix
%N is the number of observation per day
%T is the length of observation
%kn is the frequency of data
%s is the first data to use withday
%e is the last data to use withday
log_returns=diff(log_price(s:kn:e,:));
end
```

(B) Here is the graph of annualized RV with different sample frequency.

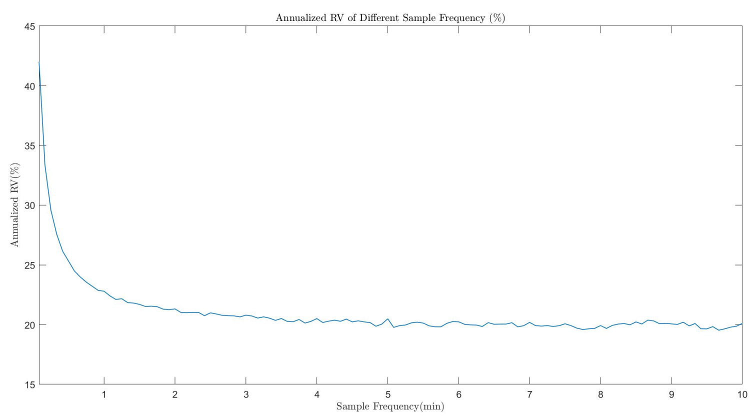

Figure 1: Annualized RV with Different Sample Frequency (%)

From the graph we can find, as the frequency decreases (k_n increases), the value the RV decreases, either. When the data frequency decreases to 2min, the value of RV becomes stable.

Since RV includes two components: $\hat{IV}$ and noise, when the data frequency is quite high, say 5-second, the value of RV computed by stock log-returns will be dominated by the noise part. As the frequency decreases, such as 5-minute, the value of RV become stable and the $\hat{IV}$ will dominate RV (the stable value of RV will be a good estimate of IV).

(C) According to part B's figure, we can find, when the data frequency is very high, the noise part will dominate RV. Motivated by this fact, the RV we calculate from high frequency data will be a good estimate of noise. In this case, $\hat{IV}$ is the "noise" of noise.

(D) Assume $E(\chi_i) = 0$ and $Cov(\chi_i, \chi_j) = 0$ ($\forall i \neq j$) and $Cov(X_{i\Delta_n}, \chi_i) = 0$, we can write the expression of RV calculated by noise log-return data as:

$$
\begin{aligned}
Y_i^n &= X_{i\Delta_n} + \chi_i \\
RV &= \sum_{i=1}^{n/k_n} (\Delta Y_{ik_n}^n)^2 \\
&= \sum_{i=1}^{n/k_n} [(X_{i\Delta_n} - X_{(i-1)\Delta_n}) + (\chi_{ik_n} - \chi_{(i-1)k_n})]^2 \\
&\approx \sum_{i=1}^{n/k_n} (\Delta_{ik_n}^n X_i)^2 + \sum_{i=1}^{n/k_n} (\chi_{ik_n} - \chi_{(i-1)k_n})^2 \\
&\approx \hat{IV} + 2\frac{n}{k_n}\hat{\sigma}_\chi^2
\end{aligned}
$$

where $\hat{IV}$ is the estimate of IV using really log-returns and $2\frac{n}{k_n}\hat{\sigma}_\chi^2$ is the noise. By the definition of $Contribution(k_n)$, we can have:

$$Contribution(k_n) \equiv \frac{2\frac{n}{k_n}\hat{\sigma}_\chi^2}{RV} = \frac{Noise}{IV + Noise}$$

(E) Here is the figure of estimated average $\widehat{Contribution}(k_n)$.

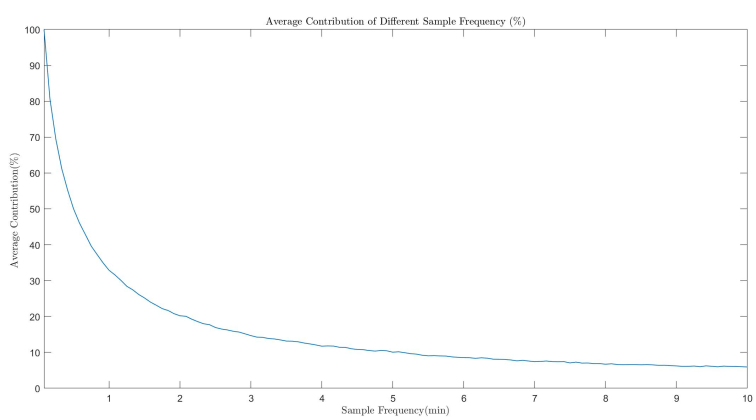

Figure 2: BAC Average Contribution of Different Sample Frequency(%)

From the figure we can see, when the data frequency is really high, the average contribution is close to 100%, which means the noise will dominate the RV; when the frequency of data decreases to 5-minute (or 8-minute), the contribution will be less than 10%, which means noise term is unimportant in estimating IV. As the data frequency keeps decreasing, the average contribution decreases further.

(F) Here is the graph of average RV and coarse sample RV based on 5-minute sample.

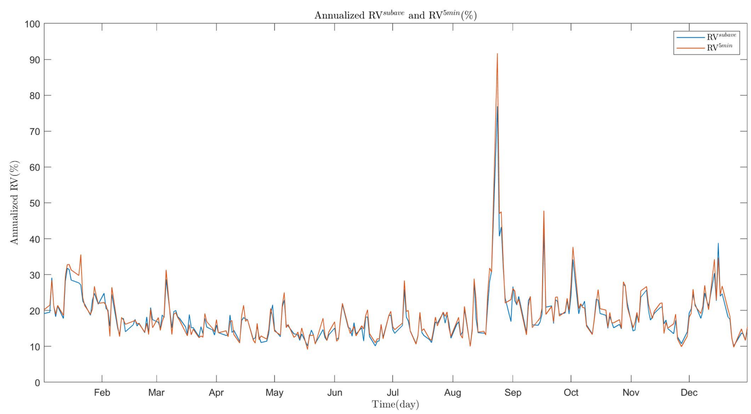

Figure 3: Annualized RV^{subave} and RV^{5min}(%)

From the figure we can see, there exists some difference between these two RV even though the difference is very small. Since we use all data to compute TSRV, Tsrv will be a better and more efficient estimator of IV compared to regular RV which only uses coarse sample data.

(G) Here is the figure of annualized TSRV and RV ^{5min}.

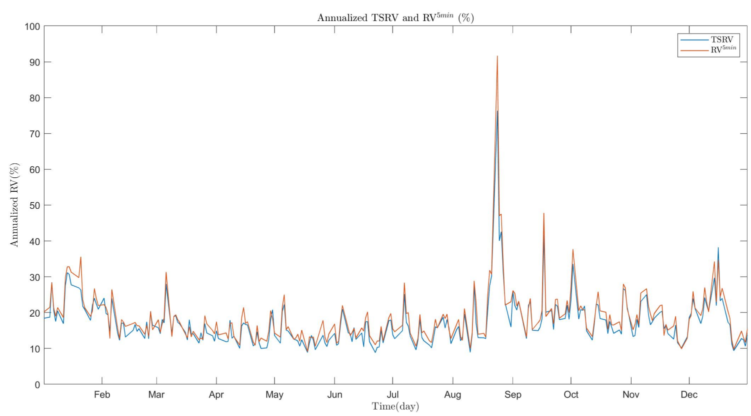

Figure 4: Annualized TSRV and RV^{5min} (%)

According the the figure we can find, the shape of TRSV is very similar to the shape of RV^{5min}, which indicates TRSV is a estimator of IV^{5min}. However,the value of TRSV is smaller than 5-min RV since we have subtracted noise term from it. After subtracting noises from RV, the value of TSRV will be much closer to IV.

(H) Here is the figure of Average Annualized TSRV and Volatility with Different Sample Frequency.

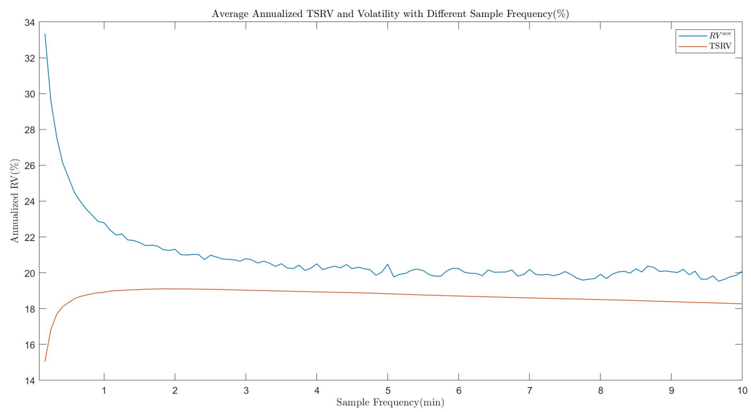

Figure 5: Average Annualized TSRV and Volatility with Different Sample Frequency(%)

For average RV, as the data frequency decreases, its decreases and becomes stable; for TSRV, whose shape is much smoother, as the data frequency decreases, TSRV decreases and becomes stable, too.

By the definition of TSRV and RV^{a}ve, the differences between these two curves are the noise. From graph we can find that, the differences of these two curve becomes smaller as the data frequency goes down, which indicates that the noise proposition in RV^{a}ve becomes smaller. This finding is consistent to the results of contribution figure.

The **MATLAB** code:

Function of Local Variance

```
addpath('D:\ZM-Documents\MATLAB\data','functions','scripts');
[dates_BAC,lp_BAC]=load_stock('BAC-2015.csv','s');
N_BAC=sum(floor(dates_BAC(1,1))==floor(dates_BAC(:,1)));% number of
    observations per day
T_BAC=size(dates_BAC,1)/N_BAC;
lp_BAC=reshape(lp_BAC,N_BAC,[]);
days_BAC=unique(floor(dates_BAC));

%B ave RV against kn
mkn=120;
RV_BAC=zeros(mkn,1);
for kn=1:mkn
  lr_BAC=log_return_new(lp_BAC,kn,1,N_BAC);
   RV_BAC(kn)=mean(sum(lr_BAC.^2));%average RV for given kn
end

figure;
plot(1/12:1/12:mkn/12,100*sqrt(252*RV_BAC));
xlabel('Sample Frequency(min)');
xlim([1/12,mkn/12]);
ylabel('Annualized RV(\%)');
title('Annualized RV of Different Sample Frequency (\%)');

%C estimate var_x using highest frequency
lr_BAC_HF=log_return_new(lp_BAC,1,1,N_BAC);
var_x_HF=(2*(N_BAC-1))^(-1)*sum(lr_BAC_HF.^2);

%E average contribution g=against varing kn
mkn=120;
cont=zeros(mkn,1);
for kn=1:mkn
  lr_BAC=log_return_new(lp_BAC,kn,1,N_BAC);
   RV_BAC=sum(lr_BAC.^2);
   %n_kn=floor((N_BAC-1)/kn);%n=(N_BAC-1)/kn
   n_kn=(N_BAC-1)/kn;%n=(N_BAC-1)/kn
   cont(kn)=100*mean(var_x_HF*(2*n_kn)./RV_BAC); %average contribution
end

figure;
plot(1/12:1/12:mkn/12,cont);
xlabel('Sample Frequency(min)');
xlim([1/12,mkn/12]);
```

```
ylim([0,100]);
ylabel('Average Contribution(\%) ');
title('Average Contribution of Different Sample Frequency (\%)');

%F average RV with different start interval j given kn=60
kn=60;
RV_BAC=zeros(kn,T_BAC);
for j=1:kn
  lr_BAC=log_return_new(lp_BAC,kn,j,N_BAC);
  RV_BAC(j,:)=sum(lr_BAC.^2);
end
RV_subave=mean(RV_BAC,1);

figure;
plot(days_BAC,100*sqrt(252*RV_subave));
hold on;
plot(days_BAC,100*sqrt(252*RV_BAC(1,:)));
xlabel('Time(day)');
xlim([min(days_BAC),max(days_BAC)]);
ylabel('Annualized RV(\%)');
title('Annualized RV$^{subave}$ and RV$^{5min}$(\%)');
datetick('x','keeplimits');
legend('RV$^{subave}$','RV$^{5min}$');

%G TSRV
%n_kn=floor((N_BAC-1)/kn); % given kn
kn=60;
n_kn=(N_BAC-1)/kn; % given kn
TSRV_BAC=RV_subave-var_x_HF*(2*n_kn);

figure;
plot(days_BAC,100*sqrt(252*TSRV_BAC));
hold on;
plot(days_BAC,100*sqrt(252*RV_BAC(1,:)));
xlabel('Time(day)');
xlim([min(days_BAC),max(days_BAC)]);
ylabel('Annualized RV(\%)');
title('Annualized TSRV and RV$^{5min}$ (\%)');
datetick('x','keeplimits');
legend('TSRV','RV$^{5min}$');

%H
%Avg TSRV agaainst different kn
% 1 given kn, calculate RVj;
```

```
% 2 then calculate subave RV;
% 3 calculate TSRV
% 4 varing kn, calculate avg TSRV
mkn=120;
TSRV_BAC=zeros(mkn,T_BAC);
RV_ave=zeros(mkn,1); %average RV
for kn=1:mkn % 4 varing kn, calculate avg TSRV
  RV_BAC=zeros(kn,T_BAC); % 1 given kn, calculate RVj;
  for j=1:kn
   lr_BAC=log_return_new(lp_BAC,kn,j,N_BAC);
   RV_BAC(j,:)=sum(lr_BAC.^2);
  end
  RV_ave(kn)=mean(RV_BAC(1,:));
  RV_subave=mean(RV_BAC,1); % 2 then calculate subave RV;
  %n_kn=floor((N_BAC-1)/kn);
  n_kn=(N_BAC-1)/kn;
  TSRV_BAC(kn,:)=RV_subave-var_x_HF*(2*n_kn); % 3 calculate TSRV
end
TSRV_ave=mean(TSRV_BAC,2);

figure;
plot(2/12:1/12:mkn/12,100*sqrt(252*RV_ave(2:end)));
hold on;
plot(2/12:1/12:mkn/12,100*sqrt(252*TSRV_ave(2:end)));
xlabel('Sample Frequency(min)');
xlim([1/12,mkn/12]);
ylabel('Annualized RV(\%)');
title('Average Annualized TSRV and Volatility with Different Sample Frequency (\%)');
legend('$RV^{ave}$','TSRV');
```