

**Boston University Questrom School of Business**  
**MF 793 – Fall 2021**  
**Take Home Final Exam**  
Eric Jacquier

**Last Name:** Liu

**First Name:** Xuyang

**ID#:** U20404938

This Take Home Final counts for a half of your total final exam grade. It is due on Gradescope before **Saturday December 11<sup>th</sup> at 12:00 pm Boston Time.**

- In Gradescope, make absolutely sure to put your answer to each question in the space provided and nowhere else otherwise it will not be graded. The space to answer each question will be clearly marked.
- You can use any documentation you wish.
- You can not communicate with anyone about the exam until the deadline has passed.
- Word answers are really a few sentences no more. If you think you need more, you misunderstanding the question. You have no time constraint so there will be no excuse for unclear or off-topic answers or answers which will get no points.
- The large number of sub-questions is to facilitate the template and clear grading in Gradescope. The Gradescope template will be posted later this week.
- We will not answer substance questions about the exam but we will answer clarifying questions.

**You have a space to put your code for all problems at the end of the exam**

### **Problem 1: Estimating Variance with and without mean!**

1) Use the daily and monthly stock files named stk-11. Use the data from 2012 to 2016 included. Use log-returns.

Estimate monthly and daily standard deviations. Report annualized values in columns (2)(4).

Estimate again, assuming negligible means, put in (3)(5)

Table 1: **Annualized** Standard Deviations with and without Mean

	Daily (2)	Daily, $\mu = 0$ (3)	Monthly (4)	Monthly, $\mu = 0$ (5)
Apple	0.2614	0.2615	0.2562	0.2580
Citigroup	0.2778	0.2778	0.2753	0.2772
Pepsi	0.1323	0.1325	0.1109	0.1153
Verizon	0.1560	0.1560	0.1637	0.1650
USVW	0.1284	0.1286	0.1048	0.1098

2) Conclude. Can we “ignore the mean” when estimating volatility for daily returns?

Yes, we can ignore the mean when estimating volatility for daily returns.

Because from what we got above, there’s no significant difference between these two daily standard deviations (with and without mean).

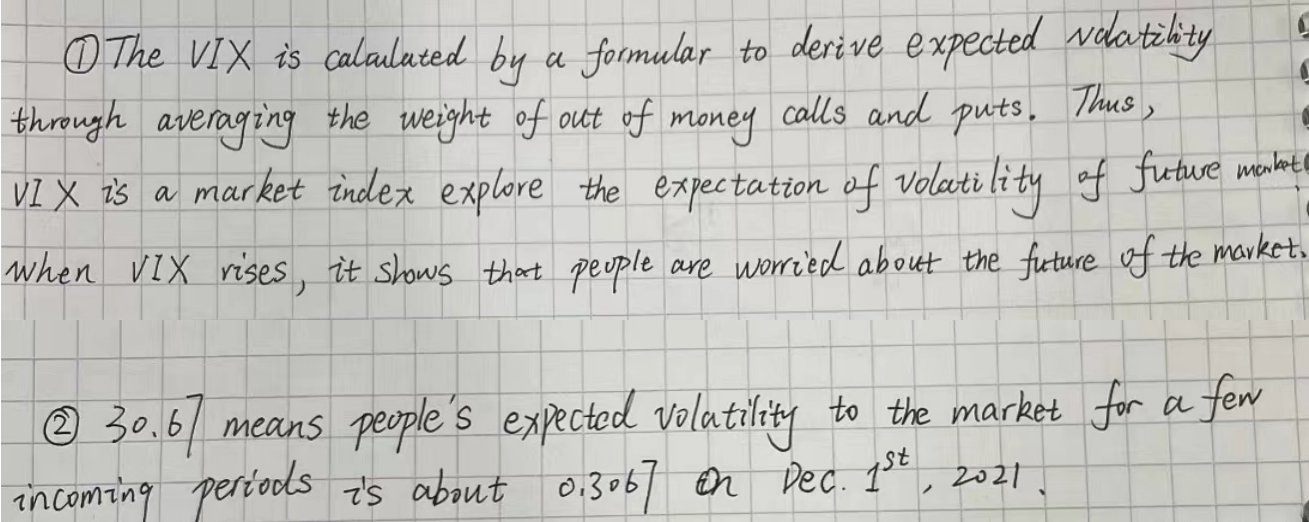
3) Conclude. For monthly returns. Is there an economic difference between estimates obtained with and without the mean?

Yes, there’s an economic difference between these estimates, because the annual std we got still have some difference.

## **Problem 2: Predicting VIX, heteroskedasticity and functional form?**

vix-2021-mon.csv contains the monthly VIX every first of the month until Dec. 1<sup>st</sup> 2021.

1) In a few hand-written lines, describe how the VIX is calculated. On Dec. 1<sup>st</sup>, 2021, the VIX was 30.67. What exactly does 30.67 estimate or forecast, for what horizon? Be specific.



① The VIX is calculated by a formula to derive expected volatility through averaging the weight of out of money calls and puts. Thus, VIX is a market index to explore the expectation of volatility of future market. When VIX rises, it shows that people are worried about the future of the market.

② 30.67 means people's expected volatility to the market for a few incoming periods is about 0.3067 on Dec. 1<sup>st</sup>, 2021.

2) This data is the monthly VIX values. Explain what you would expect to find if you plotted an ACF of the daily VIX value. Think of your answer in 1)

I would expect to see that the ACF would have the lag of a few periods. Because VIX is autocorrelated.

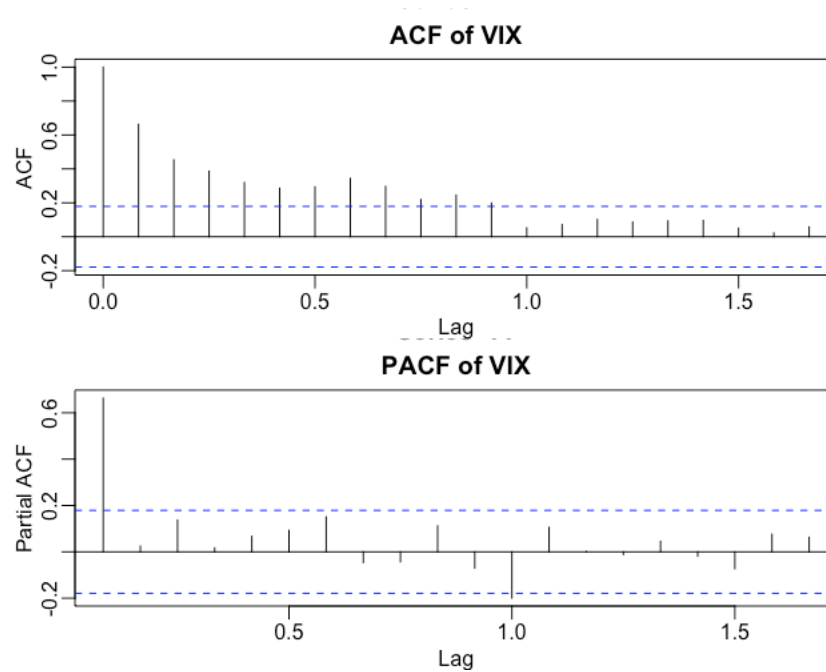
3) Select the data from 2012 to 2021, 10 years of monthly VIX. For better time series plotting, make the monthly VIX a time series, which allows to plot the months and year on the horizontal axis. Like:

```
vv<-ts(vixvalues,frequency=12,start=c(year number , month number)),  
then use ts.plot
```

Plot the time series of VIX for yourself, do not put it in your exam.

In Figure 1, put two plots, the ACF and the PACF of the monthly VIX.

Use `par(mfrow=c(2,1),mar=c(2,2,2,1),mgp=c(1.5,0.5,0))`



4) What Time Series process do you think monthly VIX follow, e.g., ARMA(?,?). Explain why.

The ACF plot tails damped sin wave, while the PACF cut off at lag 1. So this process is AR(1).

5) Use the arma command to estimate an AR(1) on VIX. **Do not use the arima command**, it has a confusing thing which we will discuss next semester. With the summary command, look at the output. Do not report but just check the output parameters for yourself.

```
myar <- arma(vix, order=c(p,q)) # an ARMA(p,q)
```

Now get the residuals. See names(myar) or names(summary(myar)).

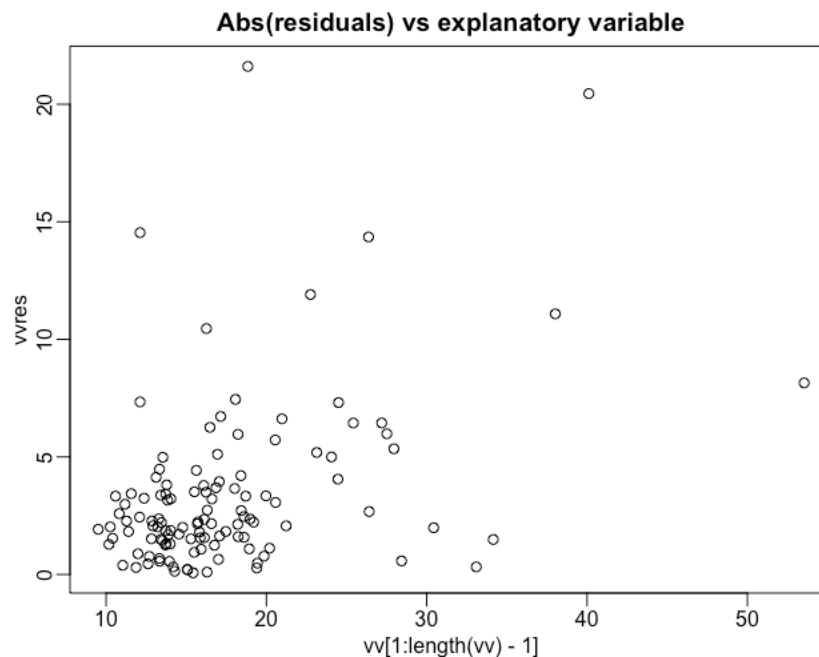
Indicate the parameter estimates and their std. dev:

	Estimate	Std.dev
Ar1	0.6858	0.0685
intercept	5.5788	1.2767

What is the first observation residual[1], why?

N/A, because it's AR(1) model, of course the first lag is absent. And also, the first wouldn't have a slope.

6) In Figure 2, plot the **absolute values** of the residuals [2:T] vs the explanatory variable vix[1:T-1]. Remember, an AR(1) is like a regression of Y on X, where X is Y at time t-1, so we try to explain vix[2:T] with vix[1:T-1].



7) What OLS assumption appears violated? How would you resolve the problem given what we discussed in Lecture Note 12 ?

It seems that the residuals still might have some relation with the variables, violating the unrelated assumption.

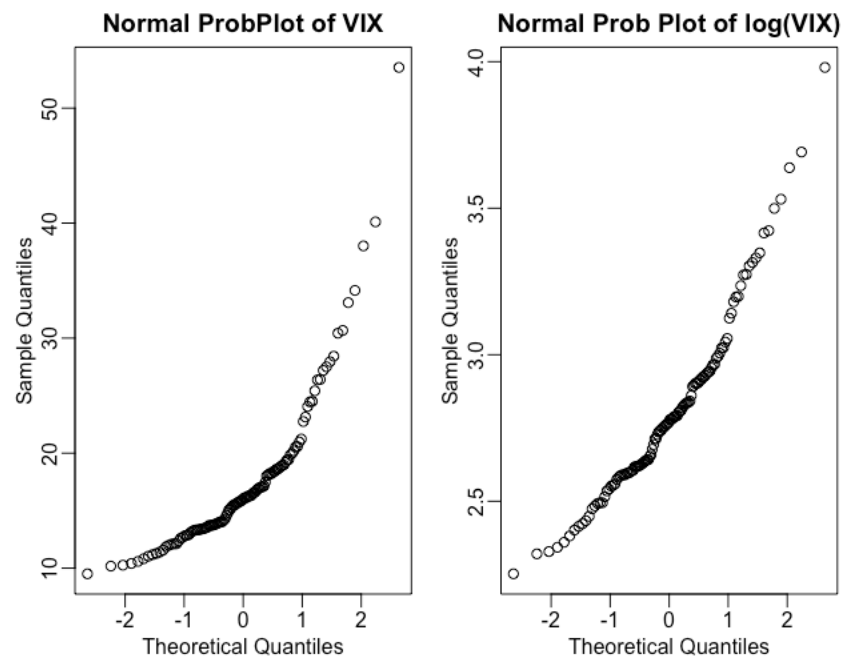
What we need to do is to add more explanation variables.

8) Don't show but just look for yourself at the normal probability plot of the residuals. From the normal probability plot, what useful assumption in the regression is violated?

The residuals are not normal contributed.

Also compute and show the Kurtosis of the residuals. **Kurt. = 12.34803.**

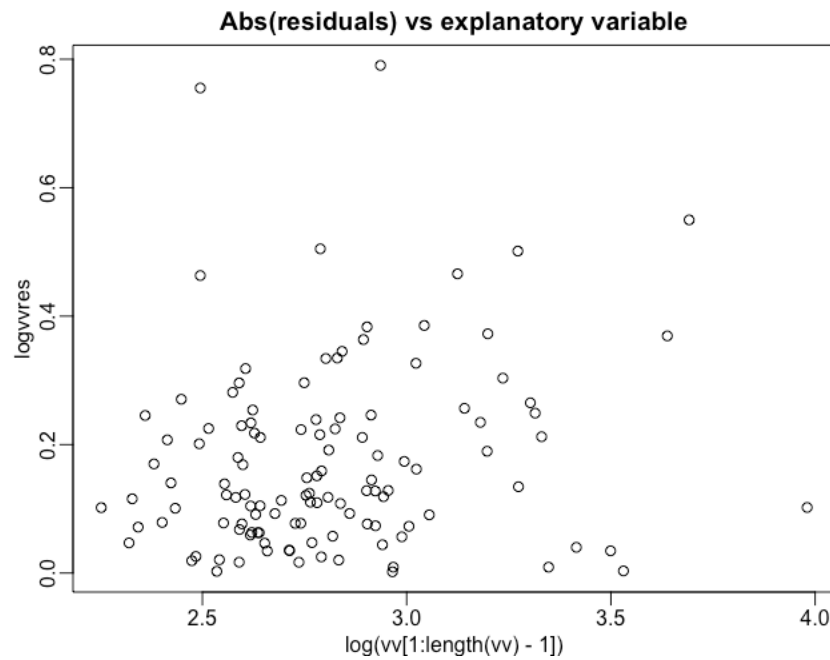
9) Let's put all this together. In Figure 3, put two plots next to each other: the normal probability plot of vix and that of log(vix).



10) Did we use the right functional form for the AR(1)? What would be a better functional form, why?

I think log form would be a better function for AR(1). As two qq-norm plot shows, log norm of VIX operates more like in a line.

**11)** Estimate an AR(1) on  $\text{Log}(\text{vix})$ , get the residuals. In Figure 4, plot the absolute values of the residuals  $\text{residual}[2:T]$  vs  $\log(\text{vix}[1:T-1])$ .



**12)** Compare Figure 4 to Figure 2. Which model is better specified, why?

Log form is better. Since figure4 tells us that there's less correlation between  $\text{abs}(\text{residuals})$  and explanatory variables.

### Problem 3: Filtering volatility

Make sure to use the R lecture notes commands like *rollapply* in the **zoo** package.

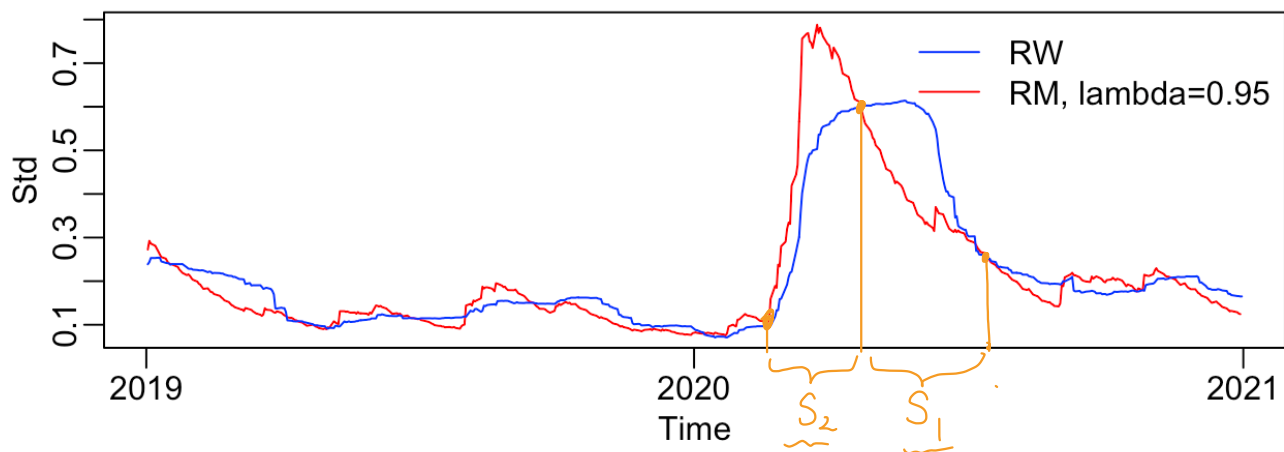
- Use the daily S&P500 return in file `spret-day-2020.csv`. Your estimators below must have a first value as of Jan. 2<sup>nd</sup> 2019, until the last day of 2020. So, your first estimates use data from 2018.
- All your estimators (RW, RM, etc..) must use the zero-mean method.

Compute a 63 day *RW* estimator of variance.

Construct a Risk-Metrics estimator (*RM*) with  $\lambda = 0.95$ . Start *RM* with an estimate of variance for Jan. 2<sup>nd</sup>, 2019 equal to the 63 day *RW* estimate on that day. Then use the updating rule.

- 1) In Figure 1, plot the annualized standard deviations produced by RW and RM, from Jan. 2<sup>nd</sup> 2019 to Dec. 31<sup>st</sup> 2020.

Figure 1: Annual Std BY RW and RM



- 2) Pointing at specific periods, explain when and why RM is superior to RW. Label these periods on your plot (you can do it by hand), explain by what criterion it is superior.

The period s1 is where Rm is superior to RW. In this period, RM is less than RW, which means the abnormal value has less influence on RM as the time go by.



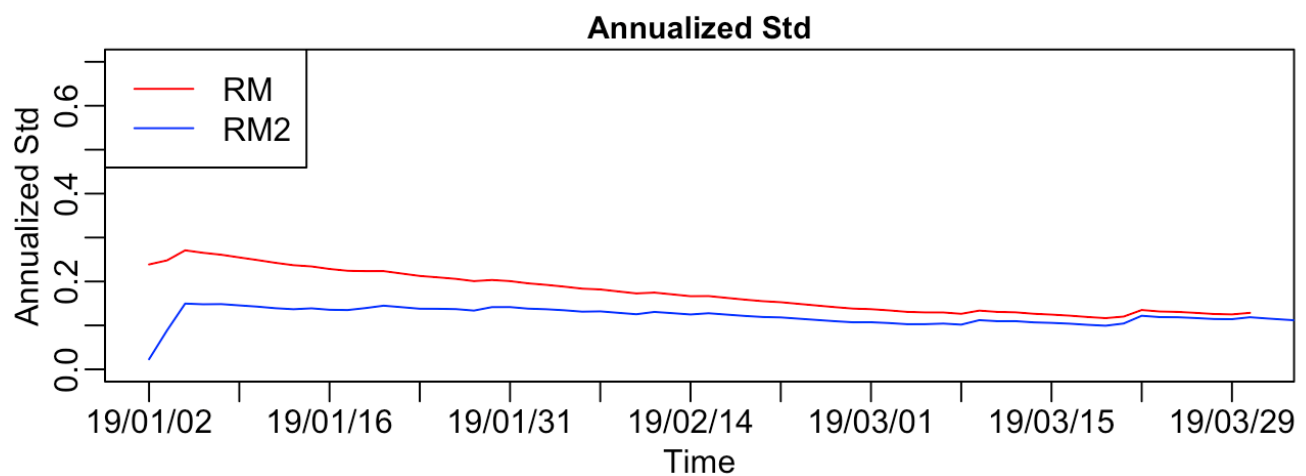
3) Similarly, explain when and how RW might be better than RM. Point and label specific periods where that might be the case on Figure 1.

During  $s_2$ , RW is better than RM. That means RW suffer less effect when abnormal value come out.

4) Now try the “other way to start” RM. For the first value on Jan. 2<sup>nd</sup>, 2019, how do you compute the “other way”? In Figure 2, plot the two versions of *RM* for the **first 3 months** (still annualized standard deviation of course). Give the date when the initial value becomes clearly irrelevant.

We start the other way with: **the square of the return on Jan. 2<sup>nd</sup> 2019**

The initial value becomes clearly irrelevant by **Mar. 29<sup>th</sup> 2019**



This is not finished .. next page is for your code

## PUT the R code for your 3 problems here:

#Problem 1

#1)

```
stkmon <- read.csv("/Users/liuxuyang/Desktop/BU\ FALL\ 2021/MF793/TakeHome/stk-11-mon-2010-2017\ .csv",header=T)
stkday <- read.csv("/Users/liuxuyang/Desktop/BU\ FALL\ 2021/MF793/TakeHome/stk-11-day-2010-2017\ .csv",header=T)
```

logmonret <-

```
cbind(stkmon[1],log(1+stkmon[2]),log(1+stkmon[5]),log(1+stkmon[8]),log(1+stkmon[,12:13]))
logdayret <- cbind(stkday[1],log(1+stkday[2]),log(1+stkday[5]),log(1+stkday[8]),log(1+stkday[,12:13]))
logmonret <- logmonret[logmonret$date>="20120101" & logmonret$date<="20161231",]
logdayret <- logdayret[logdayret$date>="20120101" & logdayret$date<="20161231",]
```

sdday <- apply(logdayret[,2:6],2,sd)

sdday <- sdday \* sqrt(252)

sdmon <- apply(logmonret[,2:6],2,sd)

sdmon <- sdmon \* sqrt(12)

mu0sdday <- apply(logdayret[,2:6]^2,2,mean)

mu0sdday <- sqrt(mu0sdday) \* sqrt(252)

mu0sdmon <- apply(logmonret[,2:6]^2,2,mean)

mu0sdmon <- sqrt(mu0sdmon) \* sqrt(12)

#Problem 2

library(forecast)

library(tseries)

#3)

```
vix <- read.csv("/Users/liuxuyang/Desktop/BU\ FALL\ 2021/MF793/TakeHome/vix2021-mon.csv",header=T)
vv <- vix[vix$date >="201201" & vix$date <="202112",][,2]
vv <- ts(vv,frequency=12,start=c(2012,1))
ts.plot(vv,ylab="VIX",xlab="Time")
```

par(mfrow=c(2,1),mar=c(2.5,2.5,2,1),mgp=c(1.5,0.5,0))

acf(vv);title("ACF of VIX")

acf(vv,type="partial");title("PACF of VIX")

#5)

vvar <- arma(vv, order=c(1,0))

summary(vvar)

vvres <- abs(residuals(vvar)[2:length(residuals(vvar))])

par(mfrow=c(1,1))

plot(vv[1:length(vv)-1],vvres,main="Abs(residuals) vs explanatory variable")

#8)

qqnorm(vvres,main="Normal Q-Q Plot for Residuals")

qqline(vvres)

Kurt <- mean((((vvres-mean(vvres))/sd(vvres))^4)

Kurt

#9)

par(mfrow=c(1,2))

qqnorm(vv,main="Normal ProbPlot of VIX")

qqnorm(log(vv),main="Normal Prob Plot of log(VIX)")

```
#11)
logvvar <- arma(log(vv), order=c(1,0))
logvvres <- abs(residuals(logvvar)[2:length(residuals(logvvar))])
par(mfrow=c(1,1))
plot(log(vv[1:length(vv)-1]),logvvres,main="Abs(residuals) vs explanatory variable")
```

```
#Problem 3
library(zoo)
winlen <- 63
spret <- read.csv("/Users/liuxuyang/Desktop/BU\ FALL\ 2021/MF793/TakeHome/spret-day-2020.csv",header=T)
azdret <- spret[spret[,1]>20181001 & spret[,1]< 20210000,2]
#RW estimator
varvec <- rollapply(azdret^2,winlen,mean)# use mean = 0 method
rw <- sqrt(varvec * 252)
```

```
#RMd
lam <- 0.95
rm <- rep(0,length(varvec))
rm[1]<- varvec[1]
for(i in 64:length(azdret)){
  rm[i-62]<-lam*rm[i-63]+(1-lam)*azdret[i]^2
}
rm<-sqrt(rm * 252)
```

```
spret <- spret[spret[,1]>20190000 & spret[,1]<20210000,c(1,2)]
zdates <- as.Date(as.character(spret[,1]),format="%Y%m%d")
par(mfrow=c(2,1),mgp=c(1.5,0.5,0))
zlength <- length(zdates)
plot(ts(rm),ylab="Annualized Std Deviation",col="blue",xaxt="n")
lines(ts(rw),col="red")
axis(1,at=seq(1,zlength-winlen+1,32),format(zdates[seq(winlen,zlength,32)],"%y/%m"))
title("RW vs RM",line=0.3,cex.main=0.95)
legend("topleft",c("RW","RM"),col=c("red","blue"),lwd=1,bty="n")
```

```
#4)
rm_2 <- rep(0,length(varvec))
rm_2[1] <- spret[spret[,1]==20190102,2]^2
for(i in 64:length(azdret)){
  rm_2[i-62]<-lam*rm_2[i-63]+(1-lam)*azdret[i]^2
}
rm_2 <- sqrt(rm_2 * 252)
```

```
rm <- rm[1:62]
rn_2 <- rm_2[1:62]
ylim=c(0,0.7)
plot(ts(rm),ylab="Annualized Std",col="red",xaxt="n",ylim=c(0,0.7))
lines(ts(rm_2),col="blue")
axis(1,at=seq(1,62,5),format(zdates[seq(1,62,5)],"%y/%m/%d"))
title("Annualized Std",line=0.3,cex.main=0.95)
legend("topleft",c("RM","RM2"),col=c("red","blue"),lwd=1,bty=1)
```