Two Sigma Data Science Challenge - Section 1

Code

Marcelo Rodrigues dos Santos

20 maio 2019

Hide

```
library(dplyr)
library(ggplot2)
library(ggpubr)
library(reshape2)
```

Hide

```
data_folder <- paste(getwd(), '/../datasets/', sep='')
```

Stock Market Data

Preparing data

Opening Dataset

Hide

```
dfretorig <- read.csv(paste(data_folder,"returns_20181228.csv",sep=''), sep=',')
dfretorig$Date<-as.POSIXct(strptime(as.character(dfretorig$Date), "%Y-%m-%d"))
dfret<-dfretorig
```

First validation was checking if weeks were completed. There is at least one week with less than 5 days registered… but it could not be an issue.

Hide

```
a<-filter(dfret,as.POSIXlt(Date)$wday==0|as.POSIXlt(Date)$wday==6)
nrow(dfret)/5
```

```
[1] 199.6
```

Analyzing the dataset I realized some stocks that only had positive returns during the studied period and higher average values. In this sense, when applying an accummulative approach to returns was causing an exponential effect. See this example.

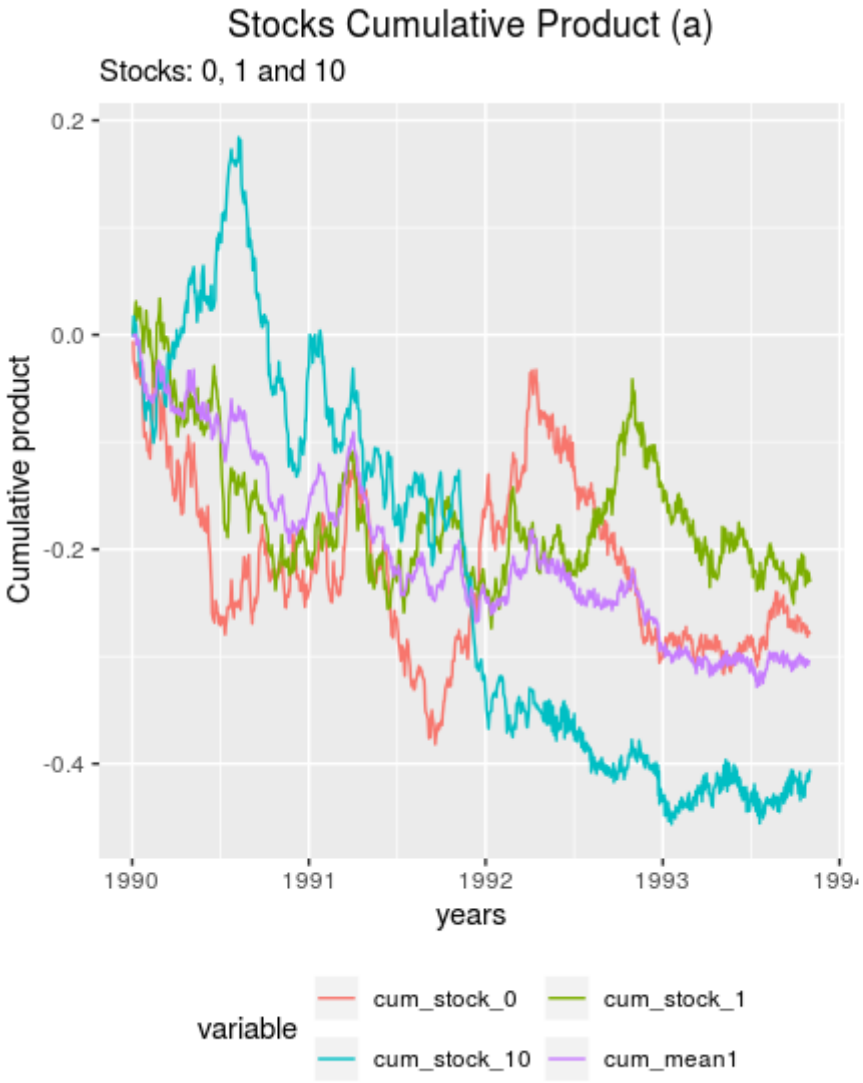
Hide

```
# stock_0
dfretaux <- dfret %>% select(Date, stock_0,stock_1,stock_10,stock_24,stock_265)%>%mutate(cum_stock_0=cumprod(1 +
 stock_0/100)-1)
dfretaux$Date<-as.POSIXct(strptime(as.character(dfretaux$Date), "%Y-%m-%d"))
# stock_1
dfretaux <- dfretaux %>% mutate(cum_stock_1=cumprod(1 + stock_1/100)-1)
# stock_10
dfretaux <- dfretaux %>% mutate(cum_stock_10=cumprod(1 + stock_10/100)-1)
# stock_24
dfretaux <- dfretaux %>% mutate(cum_stock_24=cumprod(1 + stock_24/100)-1)
# stock_265
dfretaux <- dfretaux %>% mutate(cum_stock_265=cumprod(1 + stock_265/100)-1)
```

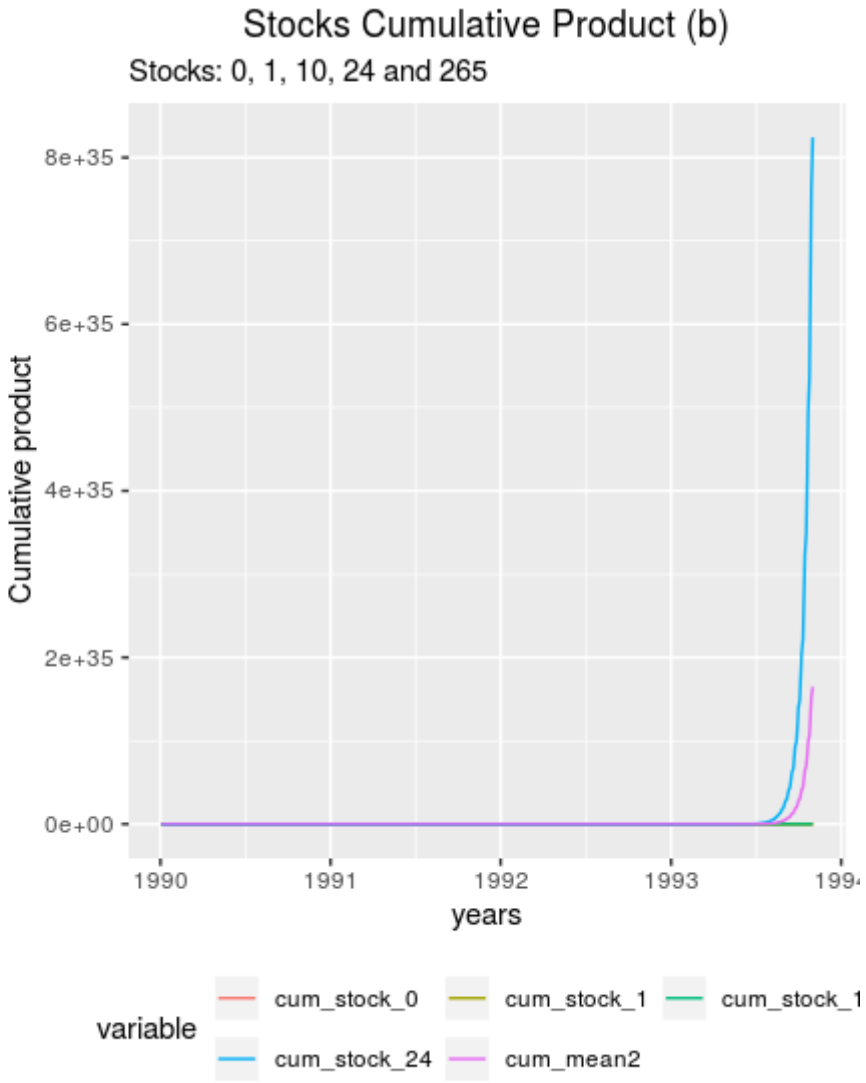
See Stocks Cumulative Product graphs.

Hide

```
dfretaux$cum_mean1<-(dfretaux$cum_stock_0+dfretaux$cum_stock_1+dfretaux$cum_stock_10)/3
a<-dfretaux[,c(1,7,8,9,12)]
dfretaux_a = melt(a, id=c("Date"))

p<-ggplot(dfretaux_a, aes(x=Date, y=value, colour=variable))+
  geom_line()+
  labs(
    title="Stocks Cumulative Product (a)", subtitle="Stocks: 0, 1 and 10", x="years", y = "Cumulative product", c
aption = "Graph 1 - Cumulative Product for stocks 0, 1 and 10") +
  theme(legend.position="bottom", plot.title = element_text(hjust = 0.5),
    plot.caption = element_text(hjust = 0.5))+
  guides(colour=guide_legend(nrow=2,byrow=TRUE))

dfretaux$cum_mean2<-(dfretaux$cum_stock_0+dfretaux$cum_stock_1+dfretaux$cum_stock_10+dfretaux$cum_stock_24+dfreta
ux$cum_stock_265)/5
a<-dfretaux[,c(1,7,8,9,10,13)]
dfretaux_a = melt(a, id=c("Date"))

q<-ggplot(dfretaux_a, aes(x=Date, y=value, colour=variable))+
  geom_line()+
  labs(
    title="Stocks Cumulative Product (b)", subtitle="Stocks: 0, 1, 10, 24 and 265", x="years", y = "Cumulative pr
oduct", caption = "Graph 2 - Cumulative Product for stocks 0, 1, 10, 24 and 265") +
  theme(legend.position="bottom", plot.title = element_text(hjust = 0.5),
    plot.caption = element_text(hjust = 0.5))+
  guides(colour=guide_legend(nrow=2,byrow=TRUE))

ggarrange(p,q,ncol = 2, nrow = 1)
```



Graph 1 - Cumulative Product for stocks 0, 1 and 10



Graph 2 - Cumulative Product for stocks 0, 1, 10, 24 and 265

Stocks with behavior presented in graph (b) have returns average ranging from 250% to 1200%. Listing all stocks with this same behavior.

Hide

```
dfret<-dfretorig
a<-data.frame()
qtde<-ncol(dfret)-1
for (i in 2:(qtde+1)) {
  a[i,'i']<-ifelse(min(dfret[,i],na.rm=TRUE)>=0,i,-1)
  a[i,'name']<-colnames(dfret)[i]
  a[i,'mean']<-mean(dfret[,i],na.rm=TRUE)
  a[i,'sd']<-sd(dfret[,i],na.rm=TRUE)
  a[i,'max']<-max(dfret[,i],na.rm=TRUE)
}
aa<-filter(a,i>=0)
head(aa)
```

	i	name	mean	sd	max
	<dbl>	<chr>	<dbl>	<dbl>	<dbl>
1	26	stock_24	8.646928	1.2672130	13.397187
2	28	stock_26	7.605875	1.3603949	11.869122
3	46	stock_44	7.720910	0.9748104	10.570228
4	49	stock_47	7.840140	1.2234774	12.538382
5	50	stock_48	6.353033	0.9394883	9.705307
6	51	stock_49	7.176961	1.0569560	11.193172

6 rows

As it is clearly a mistake… these columns (stocks) must be removed…

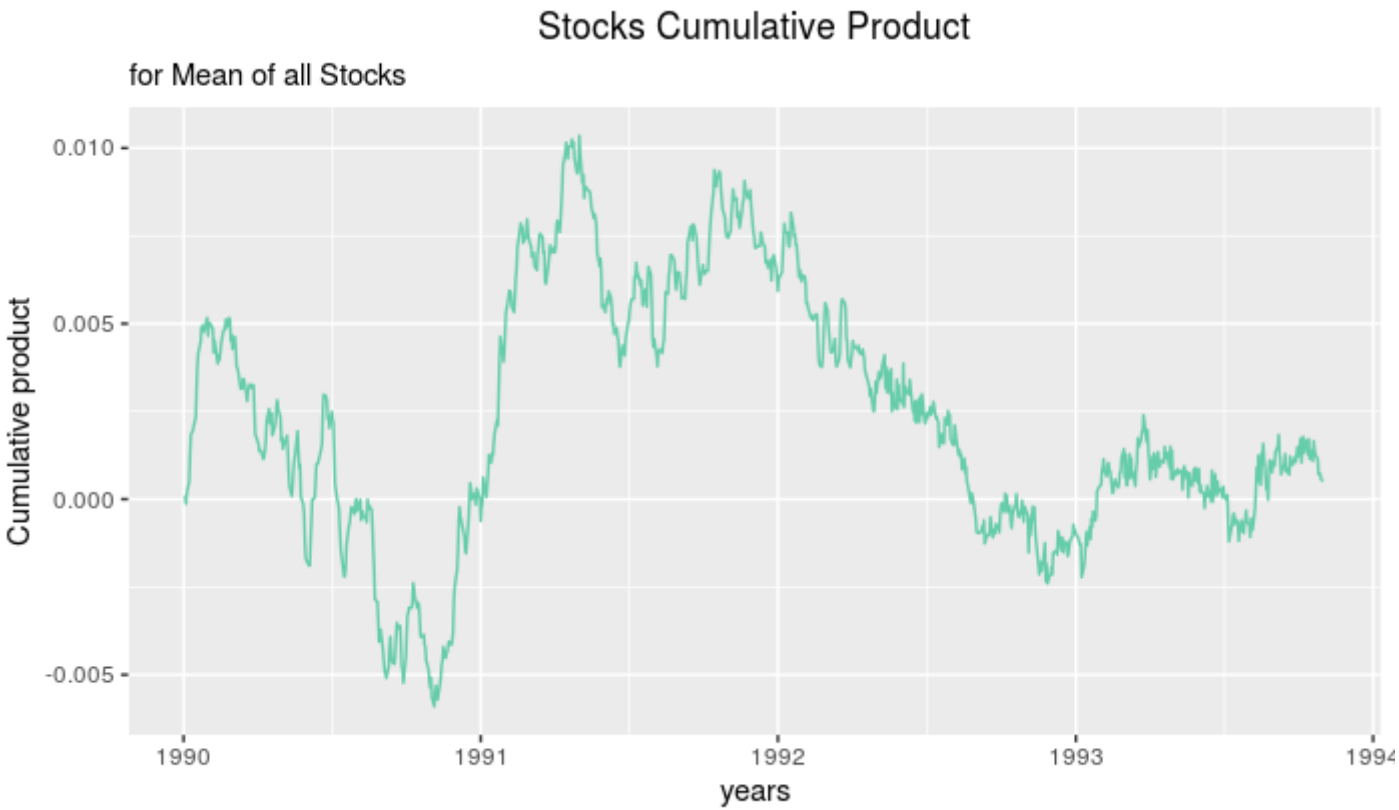
Hide

```
dfret<-dfret[,-c(aa$i)]
```

Plotting the daily return average without NAs and stocks with exponential behavior.

Hide

```
qtde<-ncol(dfret)-1
dfretnew <- dfret[,2:(qtde+1)]/100
dfretnew$media <- dfretnew[,1:qtde]%>%rowSums(na.rm=TRUE)
dfretnew$media <- dfretnew$media/qtde
dfretnew <- dfretnew %>% mutate(acumout=cumprod(1 + media)-1,acumoutsig1=lag(acumout,1))
dfretnew$Date <- dfret$Date

ggplot(dfretnew, aes(x=Date)) +
  geom_line(aes(y=acumout),colour = 'aquamarine3',na.rm=TRUE) +
  labs(
    title="Stocks Cumulative Product", subtitle="for Mean of all Stocks", x="years", y = "Cumulative product", ca
ption = "Graph 3 - Cumulative Product for Mean of all stocks") +
  theme(legend.position="bottom", plot.title = element_text(hjust = 0.5),
    plot.caption = element_text(hjust = 0.5))+
  guides(colour=guide_legend(nrow=2,byrow=TRUE))
```



Graph 3 - Cumulative Product for Mean of all stocks

Question #1:

In what month did the returns shift from exhibiting mean reversion to exhibiting momentum, or from exhibiting momentum to exhibiting mean reversion?

Generating autocorrelation probabilities for all stocks.

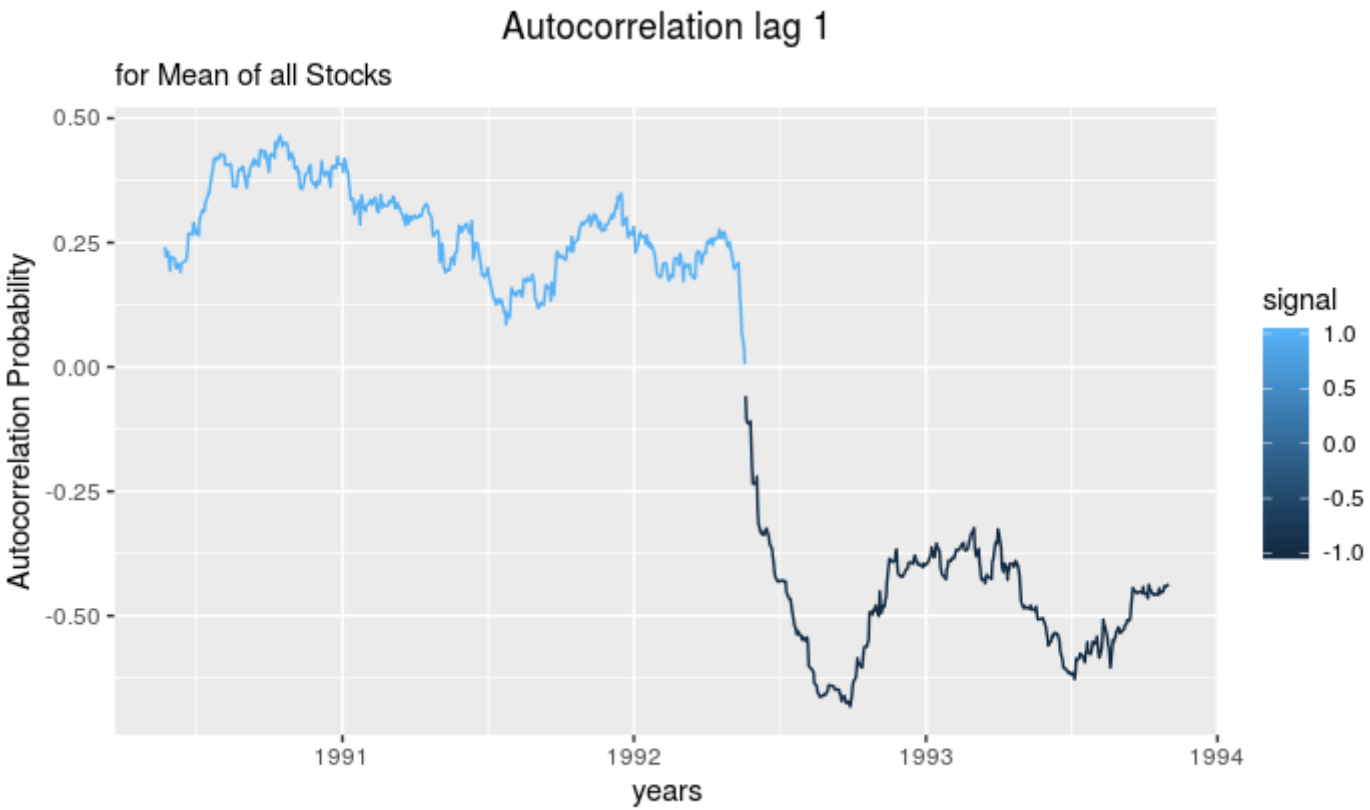
Hide

```
c<-data.frame()
qtde<-nrow(dfretnew)
for (i in nrow(dfretnew):101) {
  c[i,'acf']<-acf(dfretnew$media[(i-100):i],type='correlation',lag.max=1, plot=FALSE)$acf[2]
  c[i,'dt']<-as.POSIXct(dfretnew[i,'Date'])
  c[i,'signal']<-ifelse(c[i,'acf']>=0,1,-1)
  c[i,'media']<-dfretnew[i,'media']
}
```

Plotting daily autocorrelations

Hide

```
ggplot(c, aes(x=dt, group=signal, colour = signal)) +
  geom_line(aes(y=acf), na.rm=TRUE) +
    labs(title="Autocorrelation lag 1", subtitle="for Mean of all Stocks", x="years", y = "Autocorrelation Probab
ility",caption = "Graph 4 - Autocorrelation lag1 for Mean of all stocks") +
  theme(plot.title = element_text(hjust = 0.5),
    plot.caption = element_text(hjust = 0.5))
```



Graph 4 - Autocorrelation lag1 for Mean of all stocks

Autocorrelation starts to be negative on May 20th, 1992.

Hide

```
c[615:625,]
```

	acf <dbl>	dt <S3: POSIXct>	signal <dbl>	media <dbl>
615	0.134639803	1992-05-13	1	5.836178e-04
616	0.118097756	1992-05-14	1	-7.117023e-04
617	0.071969403	1992-05-15	1	5.962736e-04

	acf <dbl>	dt <S3: POSIXct>	signal <dbl>	media <dbl>
618	0.038805324	1992-05-18	1	-6.401806e-04
619	0.005775481	1992-05-19	1	7.072880e-04
620	-0.058420835	1992-05-20	-1	-1.204117e-03
621	-0.106265078	1992-05-21	-1	5.278600e-04
622	-0.111425612	1992-05-22	-1	-4.224977e-04
623	-0.115077757	1992-05-25	-1	-9.733431e-06
624	-0.109137687	1992-05-26	-1	8.086162e-04

1-10 of 11 rows Previous **1** 2 Next

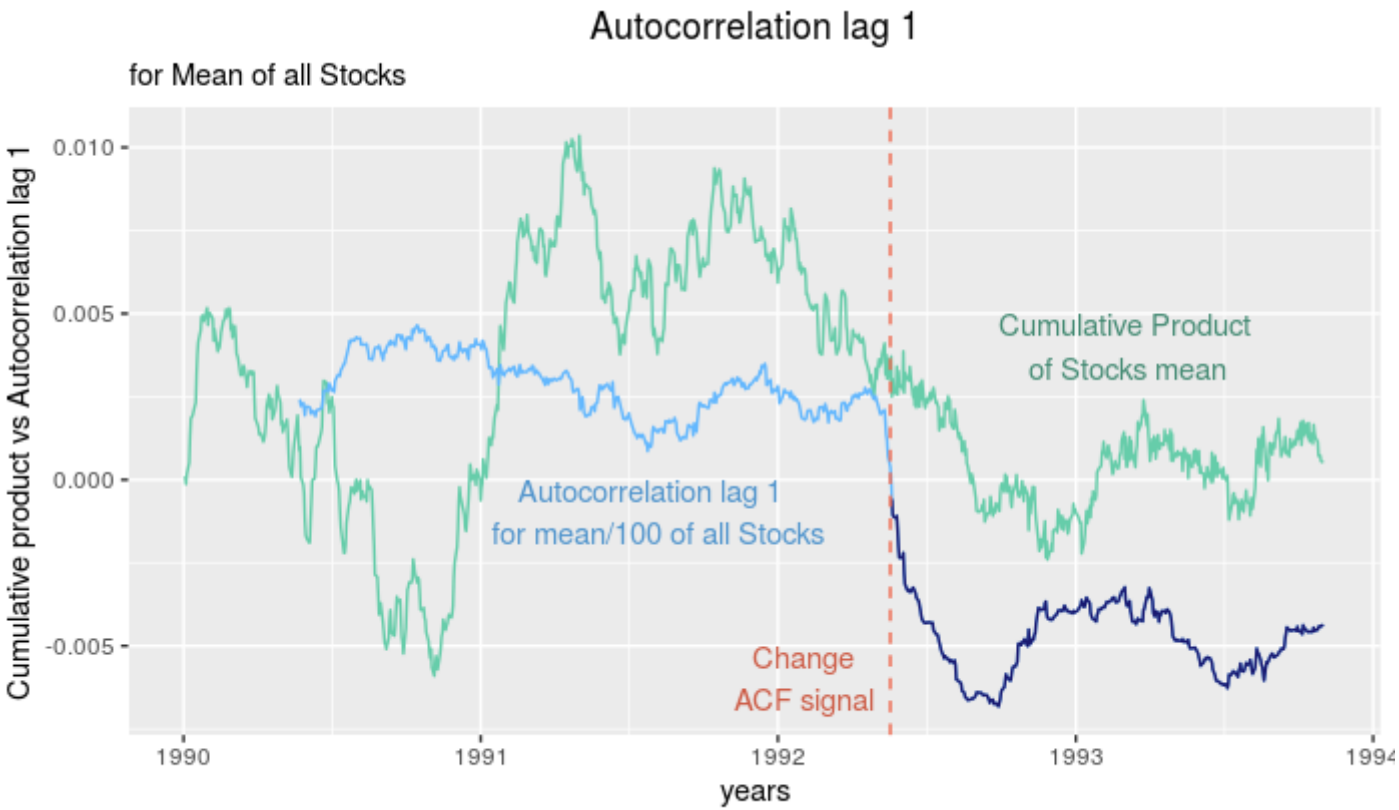
Answer for question #1 is May 1992

Question #2:

What was the average momentum? Single number: the average across all stock returns in the time period.

Hide

```
ggplot(dfretnew, aes(x=Date)) +
  geom_line(aes(y=acumout),colour = 'aquamarine3',na.rm=TRUE) +
  geom_line(aes(y=c$acf/100),colour = 'steelblue1') +
  geom_line(aes(y=c(rep_len(NA,619),c$acf[620:998]/100)),colour = 'midnightblue') +
  geom_vline(xintercept = as.POSIXct('1992-05-18'), colour='coral2',linetype = "dashed")+
  labs(title="Autocorrelation lag 1", subtitle="for Mean of all Stocks", x="years", y = "Cumulative product vs Au
tocorrelation lag 1", caption = "Graph 5 - Cumulative Product vs Autocorrelation lag 1 for Mean of all Stocks") +
  theme(plot.title = element_text(hjust = 0.5),
    plot.caption = element_text(hjust = 0.5))+
  annotate("text", x = as.POSIXct("1991-08-01"), y = -0.001, label = "Autocorrelation lag 1 \n for mean/100 of al
l Stocks", colour = "steelblue3")+
  annotate("text", x = as.POSIXct("1993-03-01"), y = 0.004, label = "Cumulative Product\n of Stocks mean", colour
= "aquamarine4")+
  annotate("text", x = as.POSIXct("1992-02-01"), y = -0.006, label = "Change\nACF signal", colour = "coral3")
```



Graph 5 - Cumulative Product vs Autocorrelation lag 1 for Mean of all Stocks

Answer for Question #2: 6.048072210^{-4}

Question #3:

What was the average mean reversion? Average during the time period when these stock returns exhibited mean reversion.

Hide

```
reversion<-subset(dfretnew,format(Date,"%Y-%m-%d")>'1992-05-19')
```

Answer for Question #3: is -8.322922710^{-4}

Question #4:

Why might the market shift from momentum being dominant to mean reversion being dominant (or the other way around)?

Answer #4: for example, if this analyzed market was a young market and then it became mature the effect of the shifting from momentum to mean reversion dominant could have happened.

Question #5:

Why might trading momentum or mean reversion succeed as a strategy? Why might it fail?

Answer #5: As we saw in Graph 5, the momentum strategy is more related to following a trend, and the mean reversion strategy means that the stock returns dynamics happen around the mean. If you are able to observe the behavior of a stock or set of assets over a certain period of time, both strategies can succeed from a mathematical and statistical point of view. However, sometimes the market may also behave differently due to unfamiliar external events and break the expected behavior or results.