

Project A: Market Data, Volatility, and Indices

Team member: Ai Zhao, Hyuk Jin Cho, Rakesh Thykandi

1A)

The Bloomberg data extract covers the data from 08/13/1980 to 09/13/2018. Our data is incomplete because we observed a lot of records that only had a single value for the entire day, compared to Yahoo which had four distinct values for open/high/low/close.

1B)

Till 04/20/1982, Bloomberg only recorded a single value for open, high, low, and close. Additionally, on 06/12/1989 Bloomberg data showed a similar behavior. These values were filtered out.

Additional data checks included

- 1) checking for negative and NA values,
- 2) checking for scenarios where High value is lower than the opening/closing/low price
- 3) checking for scenarios where Low value is greater than the opening/closing/low price.

We also found that there is a difference between previous day close and next day opening price, which can be accounted by overnight trades.

1C)

Prob ($p_{\text{high}} = p_{\text{open}}$) = 11.73%

Prob ($p_{\text{high}} = p_{\text{close}}$) = 6.59%

Prob ($p_{\text{low}} = p_{\text{open}}$) = 14.69%

Prob ($p_{\text{low}} = p_{\text{close}}$) = 1.97%

The method adopted to calculate these values involved counting the number of instances where the opening value was the highest value for the day, closing value was the highest value for the day, opening value was the lowest value for the day, and the closing value was the lowest value for the day. These values were divided by the total number of days for which we have the data to get the ratios.

Based on the calculated ratios, It is much more probable for opening price to be the high or low price compared to the closing price, which suggests that this is not a random walk event.

1D)

The top 20 intraday ranges are listed in the table below.

The period between 9/1/2008 and 8/30/2011 contains 15 occurrences.

Dates	PX_HIGH	PX_LOW	PX_OPEN	PX_LAST	range
10/19/1987	282.7	224.83	282.7	224.84	0.257394
10/20/1987	245.62	216.46	225.06	236.83	0.134713
11/13/2008	913.01	818.69	853.13	911.29	0.115208
10/10/2008	936.36	839.8	902.31	899.22	0.11498
10/28/2008	940.51	845.27	848.92	940.51	0.112674
10/9/2008	1005.25	909.19	988.42	909.92	0.105654
10/13/2008	1006.93	912.75	912.75	1003.35	0.103183
10/15/2008	994.6	903.99	994.6	907.84	0.100233
11/20/2008	820.52	747.78	805.87	752.44	0.097275
5/6/2010	1167.58	1065.79	1164.38	1128.15	0.095507

10/21/1987	259.26	236.83	236.83	258.38	0.094709
10/16/2008	947.71	865.83	909.53	946.43	0.094568
9/29/2008	1209.07	1106.39	1209.07	1106.39	0.092806
10/26/1987	248.22	227.26	248.2	227.67	0.092229
12/1/2008	888.61	815.69	888.61	816.21	0.089397
10/6/2008	1097.56	1007.97	1097.56	1056.89	0.088882
7/24/2002	844.32	775.68	797.71	843.42	0.08849
10/22/2008	951.67	875.81	951.67	896.78	0.086617
11/21/2008	801.2	741.02	755.84	800.03	0.081212
11/24/2008	865.6	801.2	801.2	851.81	0.080379

1E)

Top 20 overnight negative returns sorted reverse chronologically are shown in the table below.

The period between 1/22/2008 and 8/12/2010 contains 13 occurrences.

Dates	PX_HIGH	PX_LOW	PX_OPEN	PX_LAST	range	overnight
8/12/2010	1086.72	1076.69	1081.48	1083.61	0.009316	-0.00733
5/20/2010	1107.34	1071.58	1107.34	1071.59	0.033371	-0.00691
5/11/2009	922.99	908.68	922.99	909.24	0.015748	-0.00672
3/30/2009	809.07	779.81	809.07	787.53	0.037522	-0.00842
3/2/2009	729.57	699.7	729.57	700.82	0.04269	-0.00751
2/17/2009	818.61	789.17	818.61	789.17	0.037305	-0.00995
1/7/2009	927.45	902.37	927.45	906.65	0.027793	-0.00776
12/1/2008	888.61	815.69	888.61	816.21	0.089397	-0.00851
11/14/2008	916.88	869.88	904.36	873.29	0.05403	-0.0076
10/24/2008	896.3	852.85	895.22	876.77	0.050947	-0.01419
10/10/2008	936.36	839.8	902.31	899.22	0.11498	-0.00836
10/8/2008	1021.06	970.97	988.91	984.94	0.051588	-0.00735
1/22/2008	1322.09	1274.29	1312.94	1310.5	0.037511	-0.00924
9/25/1986	236.28	230.67	231.83	231.83	0.02432	-0.01883
7/25/1983	169.74	167.63	167.67	169.53	0.012587	-0.00722
7/5/1983	168.8	165.8	166.55	166.6	0.018094	-0.01397
10/7/1982	128.96	125.99	125.99	128.8	0.023573	-0.00772
8/27/1982	118.56	116.63	117.38	117.11	0.016548	-0.00987
6/17/1982	108.85	107.48	108.01	107.6	0.012747	-0.0079
4/29/1982	117.24	116.11	116.4	116.13	0.009732	-0.00733

Top 20 overnight positive returns sorted reverse chronologically are shown in the table below.

The period between 4/23/1982 and 8/3/1984 contains 10 occurrences.

Dates	PX_HIGH	PX_LOW	PX_OPEN	PX_LAST	range	overnight
5/10/2010	1163.85	1122.27	1122.27	1159.73	0.03705	0.010253
12/8/2008	918.57	882.71	882.71	909.7	0.040625	0.007579
11/10/2008	951.95	907.47	936.75	919.21	0.049015	0.006187
10/30/2008	963.23	928.5	939.38	954.09	0.037404	0.009988
10/14/2008	1044.31	972.07	1009.97	998.01	0.074316	0.006598

10/13/2008	1006.93	912.75	912.75	1003.35	0.103183	0.015046
9/30/2008	1168.03	1113.78	1113.78	1166.36	0.048708	0.006679
12/12/2007	1511.96	1468.23	1487.58	1486.59	0.029784	0.00672
4/3/2006	1309.19	1296.65	1302.88	1297.81	0.009671	0.006217
12/5/1988	275.62	271.81	274.93	274.93	0.014017	0.011479
8/3/1984	162.56	157.99	160.28	162.35	0.028926	0.014495
11/2/1983	165.21	163.55	165.21	164.84	0.01015	0.009594
8/17/1982	109.33	104.32	105.4	109.04	0.048025	0.012585
7/19/1982	111.78	110.66	111.75	110.73	0.010121	0.006122
7/9/1982	108.97	107.56	108.23	108.83	0.013109	0.00651
6/30/1982	111	109.5	110.95	109.61	0.013699	0.006714
6/11/1982	111.48	109.65	111.11	111.24	0.016689	0.013685
5/25/1982	115.51	114.39	115.5	114.4	0.009791	0.006185
5/6/1982	118.83	117.68	118.82	118.68	0.009772	0.009773
4/23/1982	118.64	117.19	118.02	118.64	0.012373	0.007083

1F)

The days with the top 20 jumps (as defined by the absolute ratio of the logarithmic return and the 63day trailing standard deviation is listed below). There are 3 such occurrences between 09/01/2008 and 08/30/2011.

Dates	PX_HIGH	PX_LOW	PX_OPEN	PX_LAST	range	overnight	return	sd	ratio	abs. ratio
10/19/1987	282.70	224.83	282.70	224.84	0.26	0.0000	-0.2290	0.0314	-17.6989	17.6989
10/13/1989	355.53	332.81	355.41	333.62	0.07	0.0001	-0.0632	0.0106	-9.2912	9.2912
2/5/2018	2763.39	2638.17	2741.06	2648.94	0.05	-0.0076	-0.0418	0.0076	-7.8747	7.8747
2/27/2007	1449.25	1389.42	1449.25	1399.04	0.04	-0.0001	-0.0353	0.0065	-7.6257	7.6257
10/27/1997	941.64	876.73	941.64	876.99	0.07	0.0000	-0.0711	0.0138	-6.7851	6.7851
8/8/2011	1198.48	1119.28	1198.48	1119.46	0.07	-0.0008	-0.0690	0.0140	-6.1316	6.1316
6/24/2016	2103.81	2032.57	2103.81	2037.41	0.04	-0.0045	-0.0366	0.0077	-5.9493	5.9493
8/17/1982	109.33	104.32	105.40	109.04	0.05	0.0126	0.0465	0.0099	5.9209	5.9209
8/31/1998	1033.47	957.28	1027.14	957.28	0.08	0.0000	-0.0704	0.0149	-5.8184	5.8184
2/4/1994	481.02	469.28	480.68	469.81	0.03	-0.0001	-0.0229	0.0050	-5.7390	5.7390
9/11/1986	247.06	234.67	247.06	235.18	0.05	0.0000	-0.0493	0.0112	-5.2979	5.2979
2/16/1993	444.53	433.47	444.53	433.91	0.03	-0.0001	-0.0243	0.0056	-5.2719	5.2719
8/4/2011	1260.23	1199.54	1260.23	1200.07	0.05	-0.0001	-0.0490	0.0113	-5.1273	5.1273
11/15/1991	397.16	382.62	397.15	382.62	0.04	0.0000	-0.0373	0.0083	-5.0084	5.0084
9/29/2008	1209.07	1106.39	1209.07	1106.39	0.09	-0.0032	-0.0920	0.0217	-4.9815	4.9815
2/8/2018	2685.27	2580.56	2685.01	2581.00	0.04	0.0012	-0.0383	0.0093	-4.8426	4.8426
9/17/2001	1092.54	1037.46	1092.54	1038.77	0.05	0.0000	-0.0505	0.0121	-4.8365	4.8365
8/24/2015	1965.15	1867.01	1965.15	1893.21	0.05	-0.0029	-0.0402	0.0096	-4.8124	4.8124
2/2/2018	2808.92	2759.97	2808.92	2762.13	0.02	-0.0046	-0.0214	0.0053	-4.8030	4.8030
10/16/1987	298.92	281.52	298.08	282.70	0.06	0.0000	-0.0530	0.0129	-4.7293	4.7293

2)

Data provider	Open	High	Low	Close
Bloomberg	122.00	126.97	122.00	126.97
Yahoo	122.00	125.97	122.00	125.97
NYSE	122.00	125.97	122.00	125.97

The table above shows the values by various data providers for 10/6/1982.

The recorded historical high price and close price are different. The high/close price from Bloomberg is 126.97 and the high/close price from Yahoo is 125.97. The one unit index discrepancy could correspond to a significant difference (on the order of millions of dollars) of the total market capitalization. Additionally, the closing price is the most commonly used input to measure returns of the stock market. Any errors in the closing price will have an adverse impact on the accuracy of returns which might impact investment/business decisions.

There are three scenarios: the data in Bloomberg is correct, the data in Yahoo is correct, or neither of them is correct.

We can rule out that Bloomberg's data is correct, because the difference between the opening price at 10/7/1982 and the closing price at 10/6/1982 is 1 for Bloomberg. The typical values for this difference during the time period is around 0.004, suggesting an anomaly in the Bloomberg data. The corresponding value for Yahoo is 0.01.

As a validation exercise, we cross-verified the values with a third data provider, NYSE. The values provided by NYSE match with that from Yahoo, supporting our conclusion. Thus we recommend to compare data from multiple vendors whenever such discrepancies are observed.

3)

	Date	Open	High	Low	Close	Adj. Close	Volume
Dow Jones	3/6/2015	18135.72	18135.72	17825.15	17856.78	7,856.78	113,350,000
AT&T	3/6/2015	33.58	33.77	33.37	33.48	27.89	39761500
Apple	3/6/2015	128.4	129.37	126.26	126.6	118.95	72842100
Amazon	3/6/2015	385.52	387	378.88	380.09	380.09	2627000
B. H. (A)	3/6/2015	219475	221353	218645	218811	218811	400
Divisor	0.13021608						

The table above shows the price/index information on 3/6/2015 from Yahoo Finance. The Divisor was taken from the Recitation note. We discovered that Dow Jones announced on 3/6/2015 that AT&T will be replaced by Apple with effect from 3/18/2015.

3a)

New index divisor will be **0.1354**.

- $D'(\text{Apple}) = (\sum p)/D = (17856.78 \cdot 0.13021608 - 33.48 + 126.6)/17856.78 = 0.1354$

3b)

1.44% was made up of the departing company; AT&T.

- Weight of AT&T = $p(\text{AT\&T}) / \sum p = 33.48 / (17856.78 \cdot 0.13021608) = 0.0144$

3c)

The 29 remaining companies have **greater** total weight within the index after the change from AT&T to Apple.

- Weight of remain (AT&T; before change) = $1 - p(\text{AT\&T}) / \sum p = 1 - 0.0144 = 0.9856$

- Weight of Apple = $p(\text{Apple}) / \sum p = 126.6 / (17856.78 * 0.13021608 - 33.48 + 126.6) = 0.0523$
- Weight of remain (Apple; after change) = $1 - p(\text{Apple}) / \sum p = 1 - 0.0523 = 0.9477$

3d)

New Divisor if AT&T is replaced by Amazon:

- $D'(\text{Amazon}) = (\sum p) / D = (17856.78 * 0.13021608 - 33.48 + 380.09) / 17856.78 = \mathbf{0.1496}$

New Divisor if AT&T is replaced by Berkshire Hathaway (Class A Shares):

- $D'(\text{BH A}) = (\sum p) / D = (17856.78 * 0.13021608 - 33.48 + 218811) / 17856.78 = \mathbf{12.3820}$

3e)

Visa split their share to four on the day Apple joined Dow Jones. Visa's split lowered the price of Visa which gave room for Apple to join without a major impact on the index.

Also, Apple had split their share to seven in 2014, prior to joining Dow Jones, alleviating concerns about Apple dominating the Dow Jones index.

Appendix (R Code)

```
1)

library(dplyr)

library(zoo)

#Read Bloomberg file and skip first few rows which contain unwanted information
MyData <- read.csv(file="c:/MIT/fin/sandporiginal.csv", header=TRUE, skip=4, sep=",")

#Convert to dates
MyData$Dates<-as.Date(MyData$Dates, "%m/%d/%Y")

#Remove all rows where High value = Low value
MyData<-MyData[MyData$PX_HIGH!=MyData$PX_LOW,]

# Removed any rows containing NA
MyData<-MyData[complete.cases(MyData),]

#Retained only rows having values >0
MyData<-MyData[MyData$PX_HIGH>0 & MyData$PX_LOW>0 & MyData$PX_OPEN>0 & MyData$PX_LAST>0,]

#Create boolean columns to check if High or Low values for the day occurred during opening or closing
MyData$OpenHigh<-MyData$PX_HIGH==MyData$PX_OPEN
MyData$OpenLow<-MyData$PX_LOW==MyData$PX_OPEN
MyData$CloseHigh<-MyData$PX_HIGH==MyData$PX_LAST
MyData$CloseLow<-MyData$PX_LOW==MyData$PX_LAST

#convert boolean columns to 0 or 1 to calculate averages/probabilities
logcols<-sapply(MyData,is.logical)
MyData[,logcols]<-lapply(MyData[,logcols],as.numeric)

# Calculate probability by finding the average number of times an event occurred
openhighprob<-mean(MyData$OpenHigh)
openlowprob<-mean(MyData$OpenLow)
closehighprob<-mean(MyData$CloseHigh)
closelowprob<-mean(MyData$CloseLow)

#Remove boolean columns
MyData<-MyData[,-c(6,7,8,9)]

#Define intraday range
MyData$range<-(MyData$PX_HIGH-MyData$PX_LOW)/MyData$PX_LOW

#Create new dataframe , containing dates and intra-dayrange, sorted by range
sorted<-MyData[order(-MyData$range),]

#Filter between 1/1/1980 through 8/30/2011
```

```

filtered<-sorted[sorted$Dates<"2011-08-30" & sorted$Dates>"1980-01-01",]

#Show top 20 dates
top20<-head(filtered,20)
write.csv(top20,file="1d.csv")

#Count number of instances where top20 occurrences were between 2008-09-01 and 2011-08-30
occurences<-sum(top20$Dates>"2008-09-01" & top20$Dates<"2011-08-30")

#Calculate overnight jump
MyData$overnight<-(MyData$PX_OPEN-lag(MyData$PX_LAST))/lag(MyData$PX_LAST)
MyData<-MyData[complete.cases(MyData),]

#Sort by overnight jump
Topovernight<-MyData[order(MyData$overnight),]

#Filter sorted list to cover the dates mentioned in the problem
Topovernightfiltered<-Topovernight[Topovernight$Dates<"2011-08-30" & Topovernight$Dates>"1980-01-01",]
top20positive<- tail(Topovernightfiltered,20)[rev(order(tail(Topovernightfiltered,20)$Dates)),]
top20negative<-head(Topovernightfiltered,20)[rev(order(head(Topovernightfiltered,20)$Dates)),]
write.csv(top20positive,file="1ea.csv")
write.csv(top20negative,file="1eb.csv")

```

#1F

#Calculate logarithmic returns and standard deviation of returns and ratio

```

MyData$return<-log(MyData$PX_LAST/lag(MyData$PX_LAST))
MyData$sd<-c(rep(NA,62),rollapply(data = MyData$return,width=63,FUN=sd))
MyData$ratio<-MyData$return/lag(MyData$sd)
MyData$absratio<-abs(MyData$ratio)

#Create dataframe for the top20 items by absolute value of jumps
top20jumps<-head(MyData[order(-MyData$absratio),],20)
write.csv(top20jumps,file="1f.csv")

#Count number of occurrences between 2008-09-01 and 2011-08-30 in top 20 jumps
occurences1f<-sum(top20jumps$Dates>"2008-09-01" & top20jumps$Dates<"2011-08-30")

```

2)

#Read Bloomberg file and skip first few rows which contain unwanted information

```

df_csv<- read.csv(file="~/Documents/15.458/sandp500.csv", header=TRUE, skip=4,sep=",")

#Remove all rows where High value = Low value
df_csv<-df_csv[df_csv$GSPC.High!= df_csv$GSPC.Low,]

#Removed any rows containing NA

```

```
df_csv<-df_csv [complete.cases(df_csv),]
```

```
#Retained only rows having values >0
```

```
df_csv<-df_csv [df_csv$GSPC.High>0 & df_csv$GSPC.Low >0 & df_csv$GSPC.Open >0 & df_csv$GSPC.Close >0,]
```