

Project Report

Istanbul Stock Exchange Data

Ryan Edwards

Introduction

Is there a relationship between days that are positive or negative in the stock market? Can we find patterns in the market to determine an optimal time to place a purchase order? Figuring out the answer to these questions is what can help people make the most of their money. If there are patterns that signal positive or negative days, this can be used to find optimal buy and sell points, thus maximizing the return of any investor or trader. It can even be used for retail investors who are purchasing stock in their retirement accounts. Small 1-2% fluctuations in a purchase point, ripple out into massive amounts over long periods of time. An investor looking to maximize their retirement would want the best price point. By analyzing this stock data, we can see if there are any predictions we can make with the data for the future.

Data

This section will be going over the data included in the Istanbul Stock Exchange Data. First, we will show a subset of the data and explain what is going on:

	TL BASED	USD BASED	imkb_x						
date	ISE	ISE	SP	DAX	FTSE	NIKKEI	BOVESPA	EU	EM
5-Jan-09	0.035754	0.03837619	-0.00468	0.002193	0.003894	0	0.03119	0.012698	0.028524
6-Jan-09	0.025426	0.03181274	0.007787	0.008455	0.012866	0.004162	0.01892	0.011341	0.008773
7-Jan-09	-0.028862	-0.02635297	-0.03047	-0.01783	-0.02873	0.017293	-0.0359	-0.01707	-0.02002
8-Jan-09	-0.062208	-0.0847159	0.003391	-0.01173	-0.00047	-0.04006	0.028283	-0.00556	-0.01942
9-Jan-09	0.00986	0.00965811	-0.02153	-0.01987	-0.01271	-0.00447	-0.00976	-0.01099	-0.0078
12-Jan-09	-0.029191	-0.04236116	-0.02282	-0.01353	-0.00503	-0.04904	-0.05385	-0.01245	-0.02263
13-Jan-09	0.015445	-0.00027218	0.001757	-0.01767	-0.00614	0	0.003572	-0.01222	-0.00483
14-Jan-09	-0.041168	-0.03555248	-0.03403	-0.04738	-0.05095	0.002912	-0.0403	-0.04522	-0.00868
15-Jan-09	0.000662	-0.01726784	0.001328	-0.01955	-0.01433	-0.05045	0.030314	-0.01207	-0.02343
16-Jan-09	0.022037	0.03227803	0.007533	0.006791	0.006289	0.025453	0.004867	0.008561	0.010917
19-Jan-09	-0.022692	-0.04434878	-0.05426	-0.01155	-0.00935	0.003239	-0.01315	-0.01205	-0.00403
20-Jan-09	-0.013709	-0.02966137	0	-0.01783	-0.00417	-0.02341	-0.0409	-0.01509	-0.02411
21-Jan-09	0.000865	0.00152943	0.042572	0.005011	-0.00773	-0.02056	0.033532	-0.00334	-0.00509
22-Jan-09	-0.003815	0.00504316	-0.01528	-0.00984	-0.0019	0.018818	-0.01698	-0.00655	-0.00323
23-Jan-09	0.005661	-0.01000795	0.005363	-0.00964	7.40E-05	-0.03881	0.006261	-0.00362	-0.00808
26-Jan-09	0.046831	0.06170818	0.005538	0.034787	0.037891	-0.00818	0.009838	0.0328	0.01032
27-Jan-09	-0.006635	0.01094866	0.010866	-0.0008	-0.00347	0.048148	0.004922	-0.00264	0.006344
28-Jan-09	0.034567	0.03587086	0.033007	0.044182	0.023748	0.005594	0.038725	0.029974	0.022104
29-Jan-09	-0.020528	-0.02027185	-0.03368	-0.02026	-0.02477	0.017723	-0.01475	-0.02311	0.000409
30-Jan-09	-0.008777	-0.02345826	-0.02305	-0.02048	-0.00971	-0.03167	-0.00854	-0.0072	0.002243
2-Feb-09	-0.025919	-0.03560749	-0.00053	-0.01564	-0.01745	-0.01513	-0.01629	-0.01974	-0.01909
3-Feb-09	0.015279	0.02240282	0.01571	0.02404	0.021039	-0.00617	0.027574	0.017862	0.012719
4-Feb-09	0.018578	0.02323077	-0.00752	0.026577	0.015275	0.026908	0.009565	0.01877	0.015166
5-Feb-09	-0.014133	-0.01457138	0.016233	0.003932	7.09E-05	-0.01117	0.024128	-0.00414	0.002073
6-Feb-09	0.036607	0.04275893	0.026541	0.029306	0.014788	0.015846	0.039282	0.019127	0.032338
9-Feb-09	0.011353	0.02146764	0.001484	0.004766	0.003651	-0.01341	-0.01546	0.005627	0.007895
10-Feb-09	-0.040542	-0.04390687	-0.05037	-0.03517	-0.02218	-0.0029	-0.02144	-0.02439	-0.00214

The dataset provided is a subset of the full dataset. The total set of data provides stock data from January 2009 to February 2011. This was a period where the economy was in recovery after the 2008 financial crisis. We will take note of this because this does affect how we analyze the data. During a market recovery, the trend is generally in an upward direction. This is going to make whatever analysis we do skew in the positive direction.

Each column, aside from the date column, represents the amount of percentage points that specific market moved. The date column is the day of movement that was recorded. For example, if we go to 26-Jan-09 for the NIKKEI column, we see that it moved -0.00818 or -0.818%. For ISE (Istanbul Stock Exchange), we will be using the USD BASED numbers, since the other indices track use the USD as their reference. Each index comprises all the stocks that trade in that market. There are stocks that trade in some markets, and do not trade in others. We do not have that specific data here, but the impact on our results would be minimal; there are thousands of stocks in each of the indices. Overall, the data provided in this set gives us a good foundation to work with. We can make good inferences and predictions based on this data, but we must not forget that this data will most likely be skewed in a positive direction due to the conditions of the world when this data was recorded.

Exploratory Data Analysis

We will begin looking at the data by figuring out the average returns over the period recorded for each of the indices. For each index, we get the following values:

Istanbul Stock Exchange (ISE): 0.16% average daily return (ADR)

Deutscher Aktienindex (DAX): 0.07% ADR

Standard & Poor's (SP): 0.06% ADR

Financial Times Stock Exchange (FTSE): 0.05% ADR

Nikkei heikin kabuka (NIKKEI): 0.03% ADR

Bovespa Index (BOVESPA): 0.09% ADR

(EU): 0.047% ADR

Emerging Markets (EM): 0.09% ADR

We see that the Istanbul Stock Exchange returned a daily average of 0.16% per day. Somebody investing would have made the most money over the period recorded by investing in this index. We can also look at the total return of these indices over the same period. The ISE returned 83.1%, DAX 38.6%, SP 34.4%, FTSE 27.35%, NIKKEI 16.4%, BOVESPA 50.13%, EU 25.2%, and finally EM 50.1%. As we can see here, the ISE ended up being our best performer over this period, with the closest second and third being BOVESPA and EM. This makes sense, considering Emerging Markets include Turkey and Brazil, thus they inflated the EM number. To put the amount gained in perspective, investing \$10,000 into the ISE at the first date, the investor would have an amount roughly \$18,300.

To look at the data further, we can see the number of days that ended up being positive versus negative for this period. We come to 2259 positive days, and 1919 negative days. This is expected during a recovery; generally, things are positive.

Inference

Let us check the conditions of the data to ensure is normal.

- Independence: This data is not random but could be considered random. There are 4178 pieces of data being analyzed which is less than 10% of the total days the stock markets have been active.
- Success (positive days)-Failure(negative days):
 - Success: $4178(0.5406) > 10$
 - Failure: $1919(0.4593) > 10$
- Checks are passed, \hat{p} is approximately normal.

Since our normality checks pass, we can run some other tests on this. Let us run a hypothesis test to prove that the positive days outweigh the negative days.

$H_0: p = 0.5$ (The proportion of positive days over the specified period is equal to 50%)

$H_A: p \neq 0.5$ (The proportion of positive days over the specified period is not equal to 50%)

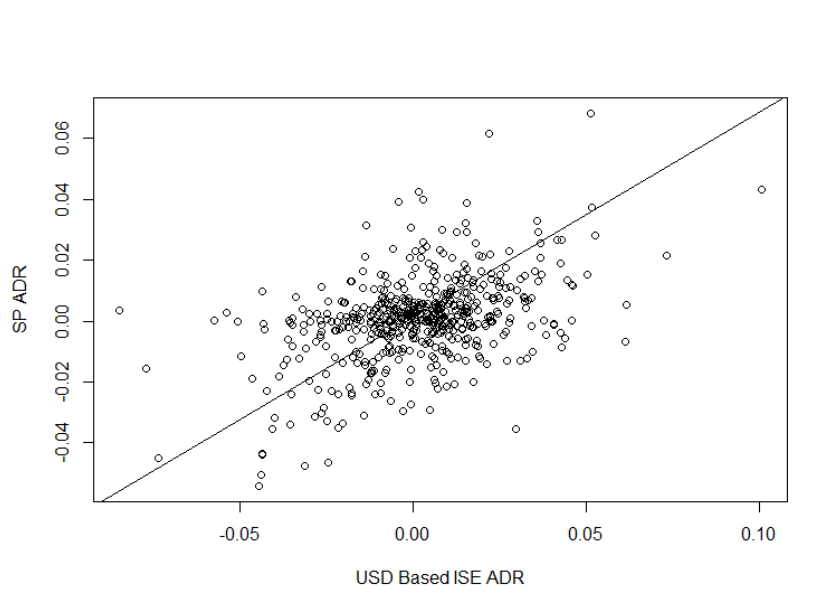
Since we have checked the conditions already, we know that the data is roughly normal. So we begin finding the test statistic:

$$Z = \frac{(\hat{p} - p_0)}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{(0.5406 - 0.5)}{\sqrt{\frac{0.5(1 - 0.5)}{4178}}} = \frac{(0.0406)}{\sqrt{0.00005984}} = \frac{(0.0406)}{0.00773563} = 5.24$$

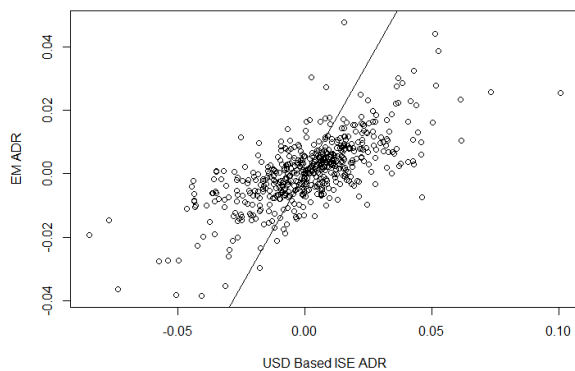
Based on this information, we fail to reject the null hypothesis that the proportion of positive days over the specified period is equal to 50%.

Regression

Let us check the regression model for this data. We will be comparing the Istanbul stock exchanges data against the index that is standard for the world, SP. When we plot the data of the USD based ISE versus the SP, we get a low correlation of 0.449, along with a scatter plot that seems to have a center.



This low correlation makes sense; the ISE is in a completely different market than SP. ISE is an emerging market, whereas SP is an American market. Let us then look at ISE versus the EM index. When we run the regression model, we get a correlation of 0.7, which makes sense since the ISE and EM are the same type of market. They *should* be correlated. The scatter plot also has a tighter shape (although the outliers skew the line of best fit by a lot).



These lines provide us with insight into how correlated different markets are with each other. We can assume that markets closer together will have higher correlations, and ones further away from one another will have lower correlations. These two tests describe this phenomenon. The only thing we can really predict based on these models, is that, in general, when markets close together (for example EU and ISE since Turkey is technically part of Europe) will *generally* move in similar directions.

By knowing that closer markets tend to move in tandem with one another, you can use this information to group different areas of the world into their own submarkets in your portfolio. Let us assume you want to own stock in some companies in the NIKKEI index. Based on our analysis, if you chose a few stocks that were geographically located in the same area, their price would generally move in the same direction. And the reverse would be true as well. This analysis would imply that Sony and Toyota, which are based in different cities in Japan, would not move in as tight a correlation as Sony and Honda would, which are based in the same section of Tokyo. All else being equal, this *should* hold true. But

since we only have the general data of the index, we cannot find this out without further analysis of the components.

Conclusion

Taking even a general look at the stock data provides an insight into how markets move, and their relation to one another. I found this project interesting; I have an interest in stocks and finance, and it was fun to be able to apply this in a school coursework setting. Based on this study, I learned that stock markets in the same geographic region move in a loose correlation with one another. I never really thought about this before doing this assignment, and it was great to discover this. I would like to apply these same concepts to test out if my theories are true. I might be able to discover some interesting correlations I did not see before. Looking at Silicon Valley companies might be an interesting experiment for the future. I would really like to take what I learned forward, and maybe see if I can answer one of my original questions that I never answered in this study: *Can we find patterns in the market to determine an optimal time to place a purchase order?* If I could find an answer to this question, which ended up being a lot more complex than I thought it would be, it could be incredibly beneficial financially. Although, I am under the impression, I am *not* the first person to have this idea. Overall, I am pleased with my findings in this project, and I hope that I can find further use in the concepts used in this assignment.

Appendix – R Code

```
# Quick Stock data analysis
library("readr")
library("plyr")
setwd("D:\\Documents\\School Documents\\Spring 2021\\Con Stat\\Research Project\\Istanbul Stock")
getwd()
stockInfo=read.csv("data_akbilgic.csv")

head(stockInfo)
attach(stockInfo)

#average daily returns
100*mean(ISE.USD)
100*mean(DAX)
100*mean(SP)
100*mean(FTSE)
100*mean(NIKKEI)
100*mean(BOVESPA)
100*mean(EU)
```

```

100*mean(EM)

#total return
100*sum(ISE.USD)
100*sum(DAX)
100*sum(SP)
100*sum(FTSE)
100*sum(NIKKEI)
100*sum(BOVESPA)
100*sum(EU)
100*sum(EM)

# Finding total positive and negative days
# with(stockInfo,count(sign(DAX)))
# we can use the individual data for later if we want

### Positives ###
posISE=nrow(stockInfo[stockInfo$ISE.USD>0,])
posDAX=nrow(stockInfo[stockInfo$DAX>0,])
posSP=nrow(stockInfo[stockInfo$SP>0,])
posFTSE=nrow(stockInfo[stockInfo$FTSE>0,])
posNIKKEI=nrow(stockInfo[stockInfo$NIKKEI>0,])
posBOVESPA=nrow(stockInfo[stockInfo$BOVESPA>0,])
posEU=nrow(stockInfo[stockInfo$EU>0,])
posEM=nrow(stockInfo[stockInfo$EM>0,])

### Negatives ###
negISE=nrow(stockInfo[stockInfo$ISE.USD<0,])
negDAX=nrow(stockInfo[stockInfo$DAX<0,])
negSP=nrow(stockInfo[stockInfo$SP<0,])
negFTSE=nrow(stockInfo[stockInfo$FTSE<0,])
negNIKKEI=nrow(stockInfo[stockInfo$NIKKEI<0,])
negBOVESPA=nrow(stockInfo[stockInfo$BOVESPA<0,])
negEU=nrow(stockInfo[stockInfo$EU<0,])
negEM=nrow(stockInfo[stockInfo$EM<0,])

# Totals
totalPos=(posISE+posDAX+posSP+posFTSE+posNIKKEI+posBOVESPA+posEU+posEM)
totalNeg=(negISE+negDAX+negSP+negFTSE+negNIKKEI+negBOVESPA+negEU+negEM)
totalPos
totalNeg

# regression
x=stockInfo$ISE.USD
y=stockInfo$SP

out=lm(x~y)

stock_plot=plot(x,y,xlab="USD Based ISE ADR",ylab="SP ADR")
stock_plot
abline(out)

# correlation of the variables
corr=cor(x,y)
corr

# confidence interval
confint(out,level=0.95)

x=stockInfo$ISE.USD
y=stockInfo$EM

out=lm(x~y)

stock_plot=plot(x,y,xlab="USD Based ISE ADR",ylab="EM ADR")
stock_plot
abline(out)

# correlation of the variables
corr=cor(x,y)
corr

```