Explaining the Stock Market with Matlab Modeling

Computer Science 170A

Nathan Tung
UCLA ID: 004059195
June 10, 2013

TABLE OF CONTENTS

INTRODUCTION

The *stock market* we have today is a network of transactions where the common individual and professional fund traders alike buy and sell company shares. Why are we so interested in exploring the stock market? First of all, the nearly-continuous exchange of shares being bought and sold at different prices yields an abundance of data for which we can data mine (that is, to parse raw data sets for patterns, visualizations, or otherwise a better understanding of the hard statistics). There is no guarantee that we will discover a trend in the fluctuation of stock prices, but the result will be nevertheless beneficial for informed investing. We will either discover a pattern that can dictate the best time to invest…or we will find that no such pattern exists and that the market is volatile and purely “random.” Secondly, in an economy where the U.S. dollar is weakening and cash placed into bonds barely if at all outpaces inflation, it makes perfect and practical sense for the average American to take an interest in stock trade. After all, every once in a while a ten-bagger (a stock that returns ten times the investment) does come along, and the stock market has allowed some of the richest people in the world to be where they are today.

All data used and conclusions reached are relevant only to the time of writing – June 2013. Using matrix-related tools such as correlation matrices, principal component analysis, (Fast) Fourier transform, polynomial regressions, and histograms of normal distributions, we will see how the stock market functions, look at the role of stock indices, and perhaps even search for indicators to predict future stock prices. Ultimately, we will look at whether the market is growing and attempt to answer for the reader: is stock market investing worth it?

STOCK MARKET INDICES

How does one determine whether the market is up (net profit) or down (net loss) for any given day? The answer generally lies in a *stock market index*, or a means of measuring the value of a portion of the market. One of these and perhaps the most well-known is the *Dow Jones Industrial Average* (DJI or DJIA), which we will simply refer to as the Dow.

The Dow tracks 30 different large public companies in various industries and, after price-weighing and scaling to account for stock splits, produces a single “average” value that represents all of its component companies. Many generalize the Dow greatly and use it to represent the entire stock market; after all, it holds some of the most important companies in the United States. Some of these well-known companies that comprise the Dow are Coca-Cola (KO), McDonald’s (MCD), Wal-Mart (WMT), Walt Disney (DIS), Microsoft (MSFT), and Bank of America (BAC). Note that there are plenty of huge corporations that are not a part of the Dow.

Other indices include National Association of Securities Dealers Automated Quotations (NASDAQ) as well as Standard and Poor’s 500 (S&P 500), the latter of which deals with a total of 500 leading companies with relatively high market capitalization in their respective industries. We will be dealing with the Dow, since it includes a smaller number of companies yet speaks volumes about the stock market.

WORTHY OF THE DOW?

So what exactly determines if a company is allowed to join the Dow? Apparently, there are no specific criteria, but editors of Wall Street Journal can choose to add big companies for the Dow 30 as long as they represent different industries. Furthermore, there seems to be no set schedule for making changes to the Dow roster. At a glance, we can tell that all the Dow companies thus far have made impressive profits and likely have a high level of *market capitalization*, or the total U.S. dollar worth of a company's shares in public circulation.

This topic is particularly interesting now that many investors are left wondering whether Apple (AAPL) or Google (GOOG), both huge tech-leading companies with plenty of cash, will soon be incorporated into the Dow. In our analysis, we will compare the current Dow 30 companies with other well-known and rather-large companies in their respective industries, with an overarching emphasis on the computer, internet, and overall technology industry. Is there any variable or pattern that distinguishes companies in the Dow and companies outside of the Dow? Are there any companies not currently in the Dow that "deserve" to be in it?

We use **principal component analysis** (PCA) on a matrix containing basic stock statistics for companies in the Dow and outside of the Dow alike. Hopefully our program will yield some kind of relationship between the "Dow-ness" (that is, either 1 for in the Dow or 0 for out of the Dow) of a stock and some of its other statistics. Some of the basic stock statistics we will be looking at include share price, average daily volume, market capitalization, earnings per share, price per sales, EBITDA (earnings before interest, taxes, depreciation, and amortization), and moving averages.

Obtaining the Data:

First, we need our raw data to fit into a matrix for principal component analysis. Yahoo! Finance is an excellent source of stock statistics, and we see that there is even a download function; we can save our file as a .csv (comma-separated values) then edit it via Microsoft Excel.

Yahoo! also allows its users to download a .csv file of stock statistics from a direct customizable URL since its default download option only provides us with transaction dates and various prices of a single share throughout the day. This is how the URL works:

http://finance.yahoo.com/d/quotes.csv?s=S&f=F

Notice that the capital S and F represent a string of characters. Replace S with many stock symbols separated by commas; replace F with tags (without separation) representing stock statistics, as shown on the following page. These are the statistics we used and their corresponding tags.

Tag (Stock Statistic)
s (symbol)
l1 (last trade price)
c1 (price change)
a2 (average volume/day)
j1 (market cap)
e (EPS)
r (P/E)
p6 (P/B)
p5 (P/S)
j4 (EBITDA)
m3 (50-day MA)
m4 (200-day MA)

Putting it all together, we obtain the URL below used to download our raw data. Keep in mind that we can use this method to find other stocks and their statistics of interest for further analysis.

http://finance.yahoo.com/d/quotes.csv?s=AA,AXP,BA,BAC,CAT,CSCO,CVX,DD,DIS,GE,HD,HPQ,IBM,INTC,JNJ,JPM,KO,MCD,MMM,MRK,MSFT,PFE,PG,T,TRV,UNH,UTX,VZ,WMT,XOM,AAPL,AMD,AMZN,ATVI,CMG,COST,DAL,DELL,FB,GOOG,HMC,LNKD,LUV,NVDA,RTN,S,STX,TGT,TM,V,WDC,YELP,YHOO&f=sl1c1a2j1erp6p5j4m3m4

After some basic editing in Excel and adding labels, we have raw data that looks like this:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Symbol	Price	Avg Vol	Market Cap	EPS	P/E	P/B	P/S	EBITDA	50-MA	200-MA	Dow?			
2	AA	8.2	17184400	8.769B	0.229	35.81	0.66	0.37	2.173B	8.4766	8.609	1	%AA		
3	AXP	74.76	5438870	82.134B	3.963	18.86	4.26	2.76	0	71.053	63.6667	1	%AXP		
4	BA	98.49	5036270	74.722B	5.327	18.49	10.01	0.92	7.795B	94.8483	82.4419	1	%BA		
5	BAC	13.09	142456992	141.1B	0.32	40.91	0.65	1.81	0	12.838	11.8409	1	%BAC		
6	CAT	84.42	7016660	55.505B	7.418	11.38	3.04	0.88	10.423B	85.9403	89.2179	1	%CAT		
7	CSCO	24.32	34975600	130.0B	1.8	13.51	2.28	2.71	13.230B	21.19	19.9894	1	%CSCO		
8	CVX	121.57	6380040	235.7B	13.232	9.19	1.68	1.08	46.203B	122.136	115.841	1	%CVX		
9	DD	54.61	6564690	50.267B	4.925	11.09	4.18	1.43	6.087B	54.2266	48.7076	1	%DD		
10	DIS	63.12	8166400	113.7B	3.301	19.12	2.7	2.6	11.353B	64.3836	56.3445	1	%DIS		
11	GE	23.32	39360200	241.1B	1.35	17.27	1.95	1.66	28.142B	22.8803	22.4181	1	%GE		
12	HD	75.1	6920520	109.7B	3.15	23.84	6.69	1.44	9.990B	75.9423	68.9508	1	%HD		
13	HPQ	24.19	22521300	46.808B	-6.852	0	1.99	0.41	13.916B	21.6639	18.6148	1	%HPQ		
14	IBM	202.74	4307820	224.8B	14.501	13.98	11.77	2.18	26.399B	202.903	201.315	1	%IBM		
15	INTC	24.7	44316000	122.8B	2.001	12.34	2.39	2.32	21.148B	21.6489	21.166	1	%INTC		
16	JNJ	83.68	9888540	235.0B	3.682	22.73	3.51	3.43	21.741B	85.7177	77.6352	1	%JNJ		
17	JPM	53.03	24147700	200.9B	5.602	9.47	1.02	2.19	0	50.5331	47.4737	1	%JPM		
18	KO	40.65	14200200	181.0B	1.911	21.27	5.57	3.78	13.148B	42.1281	39.2342	1	%KO		
19	MCD	96.42	5115680	96.666B	5.392	17.88	6.35	3.5	9.839B	100.623	95.5478	1	%MCD		
20	MMM	108.59	2807340	74.948B	6.34	17.13	4.14	2.49	7.806B	108.605	101.84	1	%MMM		
21	MRK	48.73	18354600	147.1B	1.96	24.86	2.77	3.18	16.365B	46.8503	44.2846	1	%MRK		
22	MSFT	34.78	49855600	290.5B	1.938	17.95	3.79	3.82	30.493B	28.6342	27.9396	1	%MSFT		
23	PFE	27.48	37193800	194.9B	2.089	13.15	2.37	3.38	25.335B	29.2297	27.5473	1	%PFE		
24	PG	76.66	9449100	210.1B	4.463	17.18	3.18	2.51	19.700B	78.7754	74.907	1	%PG		
25	T	35.26	25275200	189.7B	1.293	27.27	2.17	1.49	28.818B	37.0725	35.7074	1	%T		

Figure 1: Raw Stock Statistics on Excel

Using Notepad++, we are able to extract the data separated by spaces and use it in Matlab as an ordinary matrix. We remove any non-numerical characters to create the original matrix that we run PCA on. Both market cap and EBITDA are in billions of U.S. dollars; the data originally in millions were adjusted accordingly.

Matlab Code:

The code below is an example of running PCA on stock data via the correlation matrix.

```
Stocks = [
%Price, Avg Vol, Market Cap, EPS, P/E, P/B, P/S, EBITDA, 50MA, 200MA, Dow
8.2 17184400 8.769 0.229 35.81 0.66 0.37 2.173 8.4766 8.609 1; %AA
74.76 5438870 82.134 3.963 18.86 4.26 2.76 0 71.053 63.6667 1; %AXP
98.49 5036270 74.722 5.327 18.49 10.01 0.92 7.795 94.8483 82.4419 1; %A
13.09 142456992 141.1 0.32 40.91 0.65 1.81 0 12.838 11.8409 1; %AC
84.42 7016660 55.505 7.418 11.38 3.04 0.88 10.423 85.9403 89.2179 1; %CAT
24.32 34975600 130.0 1.8 13.51 2.28 2.71 13.230 21.19 19.9894 1; %CSCO
121.57 6380040 235.7 13.232 9.19 1.68 1.08 46.203 122.136 115.841 1; %CVX
54.61 6564690 50.267 4.925 11.09 4.18 1.43 6.087 54.2266 48.7076 1; %DD
63.12 8166400 113.7 3.301 19.12 2.7 2.6 11.353 64.3836 56.3445 1; %DIS
23.32 39360200 241.1 1.35 17.27 1.95 1.66 28.142 22.8803 22.4181 1; %GE
75.1 6920520 109.7 3.15 23.84 6.69 1.44 9.990 75.9423 68.9508 1; %HD
24.19 22521300 46.808 -6.852 0 1.99 0.41 13.916 21.6639 18.6148 1; %HPQ
202.74 4307820 224.8 14.501 13.98 11.77 2.18 26.399 202.903 201.315 1; %IM
24.7 44316000 122.8 2.001 12.34 2.39 2.32 21.148 21.6489 21.166 1; %INTC
83.68 9888540 235.0 3.682 22.73 3.51 3.43 21.741 85.7177 77.6352 1; %JNJ
53.03 24147700 200.9 5.602 9.47 1.02 2.19 0 50.5331 47.4737 1; %JPM
40.65 14200200 181.0 1.911 21.27 5.57 3.78 13.148 42.1281 39.2342 1; %KO
96.42 5115680 96.666 5.392 17.88 6.35 3.5 9.839 100.623 95.5478 1; %MCD
108.59 2807340 74.948 6.34 17.13 4.14 2.49 7.806 108.605 101.84 1; %MMM
48.73 18354600 147.1 1.96 24.86 2.77 3.18 16.365 46.8503 44.2846 1; %MRK
34.78 49855600 290.5 1.938 17.95 3.79 3.82 30.493 28.6342 27.9396 1; %MSFT
27.48 37193800 194.9 2.089 13.15 2.37 3.38 25.335 29.2297 27.5473 1; %PFE
76.66 9449100 210.1 4.463 17.18 3.18 2.51 19.700 78.7754 74.907 1; %PG
35.26 25275200 189.7 1.293 27.27 2.17 1.49 28.818 37.0725 35.7074 1; %T
82.24 1947490 30.957 6.586 12.49 1.21 1.21 4.513 85.1308 79.7416 1; %TRV
61.75 5992070 62.985 5.132 12.03 1.97 0.55 10.013 61.385 57.0026 1; %UNH
93.01 3297860 85.504 6.678 13.93 3.26 1.43 10.052 94.1489 89.1739 1; %UTX
48.3 12995800 138.2 0.399 121.05 4.15 1.18 31.521 51.7822 47.0965 1; %VZ
75.25 7418580 248.4 5.075 14.83 3.54 0.53 36.459 77.7626 73.0421 1; %WMT
89.65 13151000 398.6 9.83 9.12 2.39 0.96 71.864 90.1617 89.2324 1; %XOM
445.11 17939600 417.8 41.896 10.62 3.09 2.47 57.381 431.59 505.657 0; %AAPL
3.91 27123400 2.794 -0.995 0 6.73 0.57 -0.170 3.5017 3.1414 0; %AMD
267.17 3442660 121.6 -0.192 0 14.42 1.9 2.710 264.631 255.814 0; %AMZN
14.3 8606650 15.981 1.075 13.3 1.38 3.19 1.670 14.5366 12.421 0; %ATVI
353.03 497142 10.912 9.228 38.26 8.48 3.87 0.5658 365.011 350.198 0; %CMG
109.17 2032310 47.651 4.623 23.61 4.55 0.45 3.991 104.607 100.956 0; %COST
17.56 13706500 14.891 1.048 16.76 0 0.41 4.062 17.5749 16.8805 0; %DAL
13.43 32552600 23.476 1.058 12.69 2.19 0.41 3.711 14.138 11.9231 0; %DELL
22.899 45116100 55.368 0.046 497.8 4.67 10.09 1.310 26.6331 26.1612 0; %F
859.7 2320560 285.2 33.422 25.72 3.77 5.33 16.813 805.294 744.272 0; %GOOG
36.79 578934 66.307 2.589 14.21 1.04 0.53 14.426 39.9136 37.7497 0; %HMC
163.49 2308510 18.043 0.258 633.68 18.17 16.28 0.1495 178.874 177.656 0; %LNKD
13.63 7446560 9.845 0.512 26.62 1.41 0.57 1.707 13.8926 11.9943 0; %LUV
14.16 10678000 8.183 0.931 15.21 1.7 1.9 0.8897 12.6326 12.4229 0; %NVDA
65.5 1853210 21.262 5.702 11.49 2.56 0.87 3.340 63.1408 58.3466 0; %RTN
7.2 56305100 21.730 -1.366 0 3.35 0.61 5.315 7.2217 6.8368 0; %S
43.69 5713620 15.665 6.303 6.93 4.74 1.02 3.590 35.1666 31.7673 0; %STX
70.17 4461850 45.028 4.256 16.49 2.72 0.62 7.643 69.9091 65.2115 0; %TGT
114.87 671938 181.9 7.72 14.88 1.18 0.65 30.827 118.688 102.946 0; %TM
177.22 3081590 115.2 3.564 49.73 4.29 10.38 7.042 175.275 161.616 0; %V
62.81 2978300 14.847 7.931 7.92 1.77 0.91 3.836 49.9369 43.0585 0; %WDC
28.2 1416650 1.817 -0.227 0 10.96 11.63 -0.0045 28.716 23.4061 0; %YELP
25.75 19962600 27.878 3.446 7.47 1.98 5.68 1.327 23.268 20.0919 0; %YHOO
];

Stat = {'Price' 'Avg Vol' 'Market Cap' 'EPS' 'P/E' 'P/B' 'P/S' 'EBITDA' '50MA' '200MA' 'Dow'}

Name = {'AA' 'AXP' 'BA' 'BAC' 'CAT' 'CSCO' 'CVX' 'DD' 'DIS' 'GE' 'HD' 'HPQ' 'IBM' 'INTC' 'JNJ'
'JPM' 'KO' 'MCD' 'MMM' 'MRK' 'MSFT' 'PFE' 'PG' 'T' 'TRV' 'UNH' 'UTX' 'VZ' 'WMT' 'XOM' 'AAPL'
'AMD' 'AMZN' 'ATVI' 'CMG' 'COST' 'DAL' 'DELL' 'FB' 'GOOG' 'HMC' 'LNKD' 'LUV' 'NVDA' 'RTN' 'S'
'STX' 'TGT' 'TM' 'V' 'WDC' 'YELP' 'YHOO' };
```

```
n = size(Stocks,1);
p = size(Stocks,2);
StockStats = Stocks(:, 1:(p-1));
Downess = Stocks(:,p);

A = StockStats; %Ax=b
b = Downess;
numStocks = sum(Stocks(:,p))

% Covariance/Correlation Matrix analysis
CovarianceMatrix = cov(StockStats);
CorrelationMatrix = corr(StockStats);

[U,S,V] = svd(CorrelationMatrix);

Corr2PC = U(:,1:2);
XYcoordinates = StockStats * Corr2PC;
X = XYcoordinates(:,1);
Y = XYcoordinates(:,2);

figure
plot( X, Y, 'b.' )
hold on
plot( X(1:numStocks), Y(1:numStocks), 'r.' )
rotate3d on
title('Correlation PCA of Stocks (Dow stocks are red, non-Dow stocks are blue)')
xlabel('1st Principal Component')
ylabel('2nd Principal Component')

for i=1:n
   text( X(i) * 0.99, Y(i) * 1.01, Name(i) )
end
hold off

Corr2PC
```

In our PCA, we tried using both the covariance and correlation matrix of the stock data. In some cases, using the correlation matrix normalizes the values and prevents discrepancies made by scaling. However, the data between different stocks here is actually not too dissimilar in terms of magnitude, so using the covariance matrix should be sufficient here.

Results:

The covariance matrix, after undergoing singular value decomposition, produces these first two principal components below.

```
Cov2PC =
    -0.0000   -0.5795
     1.0000   -0.0000
     0.0000   -0.2408
    -0.0000   -0.0269
     0.0000   -0.0217
    -0.0000   -0.0032
    -0.0000   -0.0026
    -0.0000   -0.0201
    -0.0000   -0.5541
    -0.0000   -0.5454
```

The first column corresponds to a stock's **average daily volume**, whereas the second relates to the **stock's last trade price**. We can say that the first principal component is the average volume and the second principal component is the stock price.

Our program outputs the plot below for us to analyze.

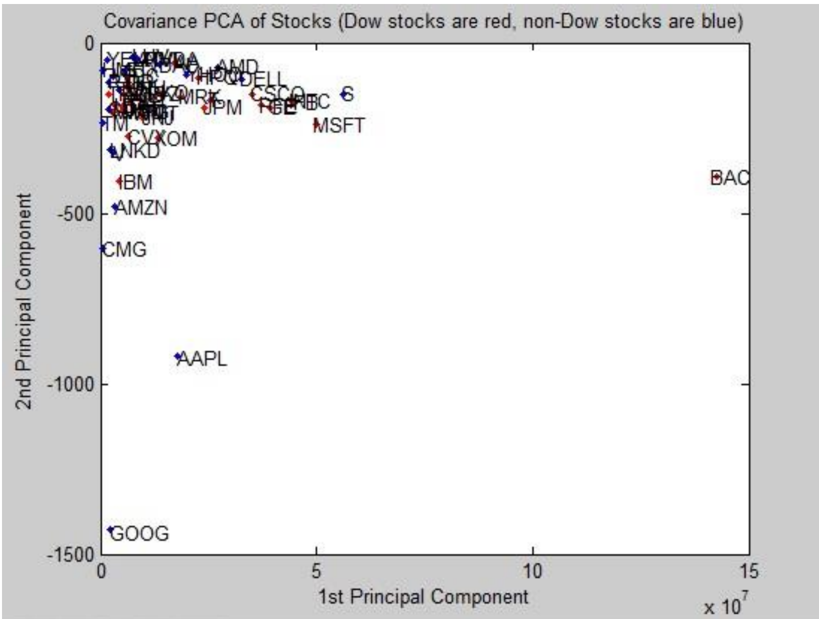

Figure 2: Stock PCA Output (Zoomed-Out)

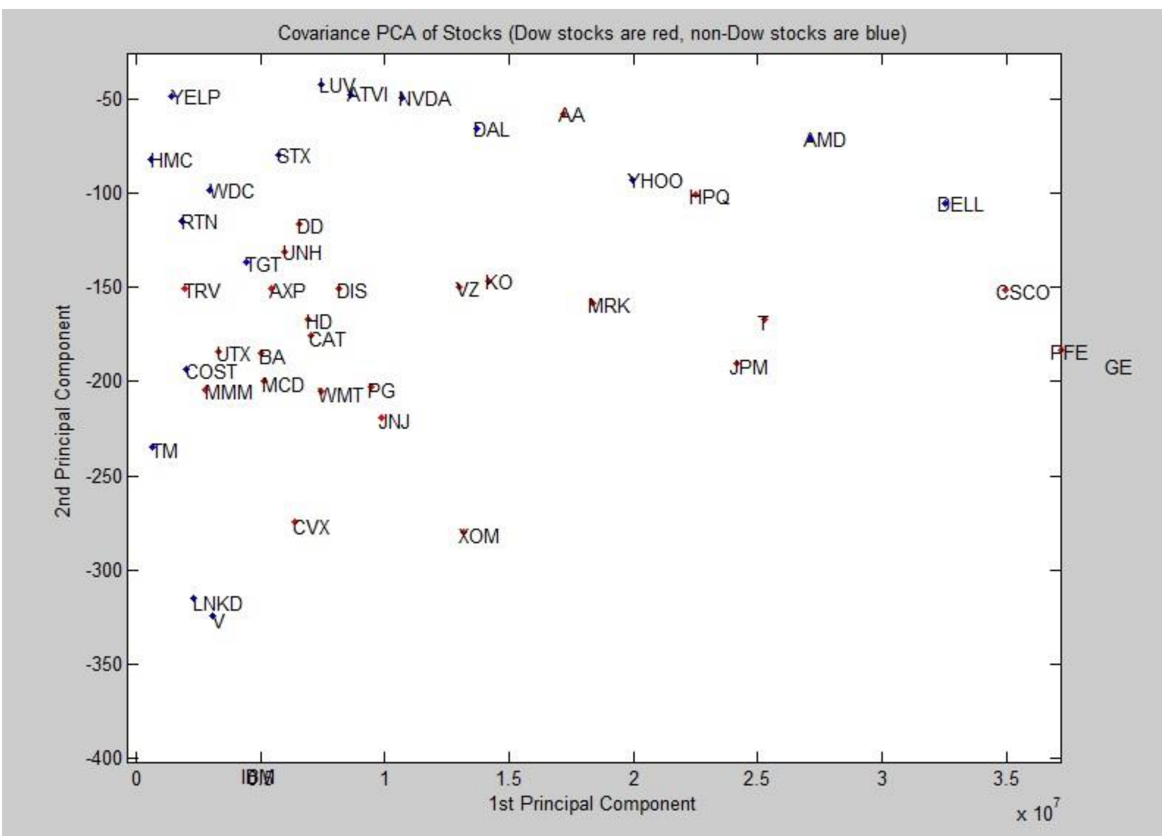

Figure 3: Stock PCA Output (Zoomed-In)

The first graph shows us that most of the stocks are clumped together in terms of daily volume and share price. In Figure 3, we enlarge the graph to better see how the Dow companies rank in terms of the two principal components.

Average Daily Volume: On the x-axis, we examine the stock's volume – that is, the average number of times that company's stock is traded in a day. It seems that the companies with the highest average daily volume are usually on the Dow, such as Bank of America (BAC), Microsoft (MSFT), Intel (INTC), etc. *Non-Dow companies that rank in high volume include Sprint (S), Facebook (FB), and Dell (DELL).*

Share Price: On the y-axis, we see the last trade price for a single share (along with some contribution from the 50- and 200-day moving average price). Interestingly, all the red (Dow) companies are mostly conglomerated within a range. The highest priced and lowest priced shares usually are non-Dow companies, whereas the Dow companies are concentrated between a second principal component of -210 and -150. *Non-Dow companies that are within this range include Costco (COST), Target (TGT), Facebook (FB), and Sprint (S).*

We confirm this trend on our Excel spreadsheet when we sort by volume and price and check the "Dow-ness." The most-traded stocks volume-wise are mostly in the Dow; in addition, most stocks within a medium-high price range are in the Dow. Though companies outside of the Dow can be in close proximity just as some Dow companies are outliers, this pattern seems strong.

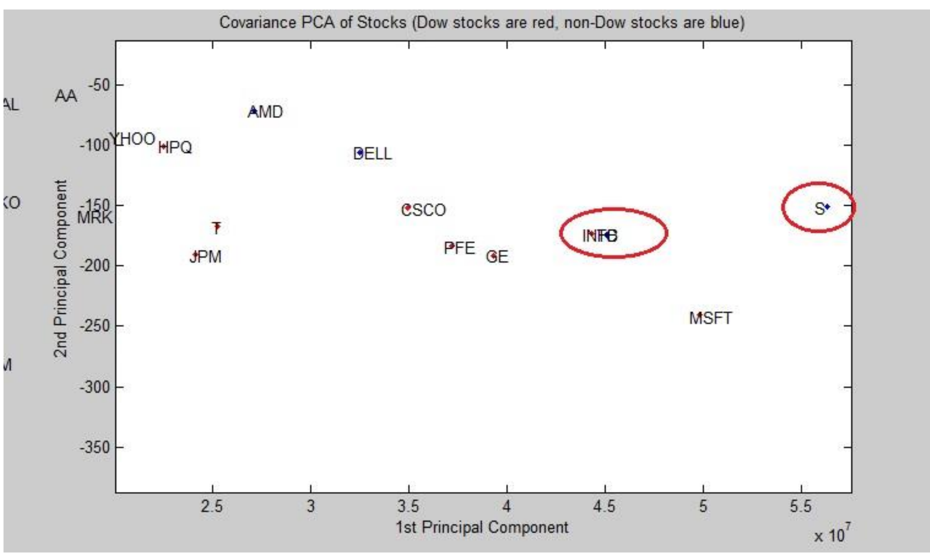

Figure 4: Facebook and Sprint are "Dow-worthy" on our PCA plot

By looking at these two PCA "Dow indicators" that we discovered, we conclude that the companies **Facebook (FB)** and **Sprint (S)** can be considered "worthy of the Dow." Of course, our conclusion is based on a select amount of stocks and assumes that Wall Street Journal selects new companies for the Dow similar to companies currently in the Dow. In reality, it is probably more likely for a large company like Apple or Google to join the Dow. But in the end, we have shown that Dow companies are all related in their high average daily volumes and their medium-to-high stock prices.

MARKET MOVEMENT

Now that we have explained the basics of stock indices like the Dow Jones Industrial Average, it is time to ask whether investing in the stock market is even worth it. Does the market actually grow over time? The answer seems obvious when we compare the valuation of indices like the Dow today with their prices in 1928 or even ten years ago. In fact, historical data shows us that the Dow was priced at a meager 250 in 1928. It grew to a value of 10,000 in the year 2000 and has recently hit an all-time high of over 15,000 points.

But knowing that the market has grown over such long periods does little to assure the investor that the stock market has more potential than other means of investments. Most people don't have the luxury of decades to hold onto their stocks. It would be far more beneficial to see how the market changes daily and analyze these minute changes. We will determine the daily price changes of the Dow and determine how its fluctuations are distributed: is it normal? Does the price change appear to be random? Or are they skewed in such a way that the market is much more likely to increase over time?

<u>Obtaining the Data:</u>

As before, we pull all the historical prices of the Dow using Yahoo! Finance's .csv download. The raw data can be downloaded using the following URL:
http://ichart.finance.yahoo.com/table.csv?s=DJIA.
Our data will be constructed by taking each day's "adjusted close" share price and subtracting the share price of the day before. The values we obtain are essentially how much the Dow has changed in price every day. This data will be used to create a histogram, which in turn can tell us about the price change distribution.

	A	B	C	D	E	F	G	H
1	Date	Open	High	Low	Close	Volume	Adj Close	Price Change
2	6/4/2013	15255.22	15354.16	15073.96	15177.54	3653840000	15177.54	-76.49
3	6/3/2013	15123.55	15297.11	15054.02	15254.03	3952070000	15254.03	138.46
4	5/31/2013	15322.22	15424.29	15112.8	15115.57	4099600000	15115.57	-208.96
5	5/30/2013	15302.8	15461.83	15220.55	15324.53	3498620000	15324.53	21.73
6	5/29/2013	15399.94	15419.64	15188.26	15302.8	3587140000	15302.8	-106.59
7	5/28/2013	15307.33	15552.39	15307.33	15409.39	3457400000	15409.39	106.29
8	5/24/2013	15290.74	15350.91	15163.64	15303.1	2758080000	15303.1	8.6
9	5/23/2013	15300.57	15378.16	15127.03	15294.5	3945510000	15294.5	-12.67
10	5/22/2013	15387.12	15562.46	15243.84	15307.17	4361020000	15307.17	-80.41
11	5/21/2013	15334.97	15481.57	15271.7	15387.58	3513560000	15387.58	52.3
12	5/20/2013	15348.33	15422.6	15257.59	15335.28	3275080000	15335.28	-19.12
13	5/17/2013	15234.75	15408	15183.26	15354.4	3440710000	15354.4	121.18
14	5/16/2013	15273.92	15355.38	15169.75	15233.22	3513130000	15233.22	-42.47
15	5/15/2013	15211.87	15325.49	15129.8	15275.69	3657440000	15275.69	60.44
16	5/14/2013	15092.15	15247	15054.22	15215.25	3457790000	15215.25	123.57
17	5/13/2013	15113.42	15159.76	15007.61	15091.68	2910600000	15091.68	-26.81
18	5/10/2013	15082.62	15162.84	15000.88	15118.49	3086470000	15118.49	35.87
19	5/9/2013	15105.12	15144.83	15046.87	15082.62	3457400000	15082.62	-22.5
20	5/8/2013	15056.2	15146.59	14957.8	15105.12	3554700000	15105.12	48.92
21	5/7/2013	14968.89	15100.89	14928.16	15056.2	3309580000	15056.2	87.31
22	5/6/2013	14973.96	15056.28	14882.13	14968.89	3062240000	14968.89	-5.07
23	5/3/2013	14838.34	15066.8	14838.34	14973.96	3603910000	14973.96	142.38
24	5/2/2013	14700.95	14880.46	14682.13	14831.58	3366950000	14831.58	130.63
25	5/1/2013	14839.8	14871.51	14650.68	14700.95	3530320000	14700.95	-138.85

Figure 5: Dow Jones Price Changes on Excel

Matlab Code:

```
PriceChange=[…]; %price change in the Dow for each day
nbins = 20; %specify how many column increments we want
hist(PriceChange,nbins);

[bincount,bincenter]=hist(PriceChange,nbins);
n=sum(bincount);
stepsize=bincenter(2)-bincenter(1);
scale=n*stepsize;
xbar=mean(PriceChange)
s=std(PriceChange)
x=bincenter;
hold on
gaussian_fit = 1/(sqrt(2*pi)*s) * exp( -1/2 * ((x-xbar)/s).^2 );
plot(x, scale*gaussian_fit, 'r');
hold off
```

Our code here creates a histogram of all the Dow price change data with `nbins=20` number of columns. We output the mean and standard deviation of the daily price change. Lastly, a Gaussian fit is placed over the histogram plot to demonstrate what a normal curve might look like in our situation.

Results for 1928-2013:

```
xbar = 0.7138
s = 54.5409
```

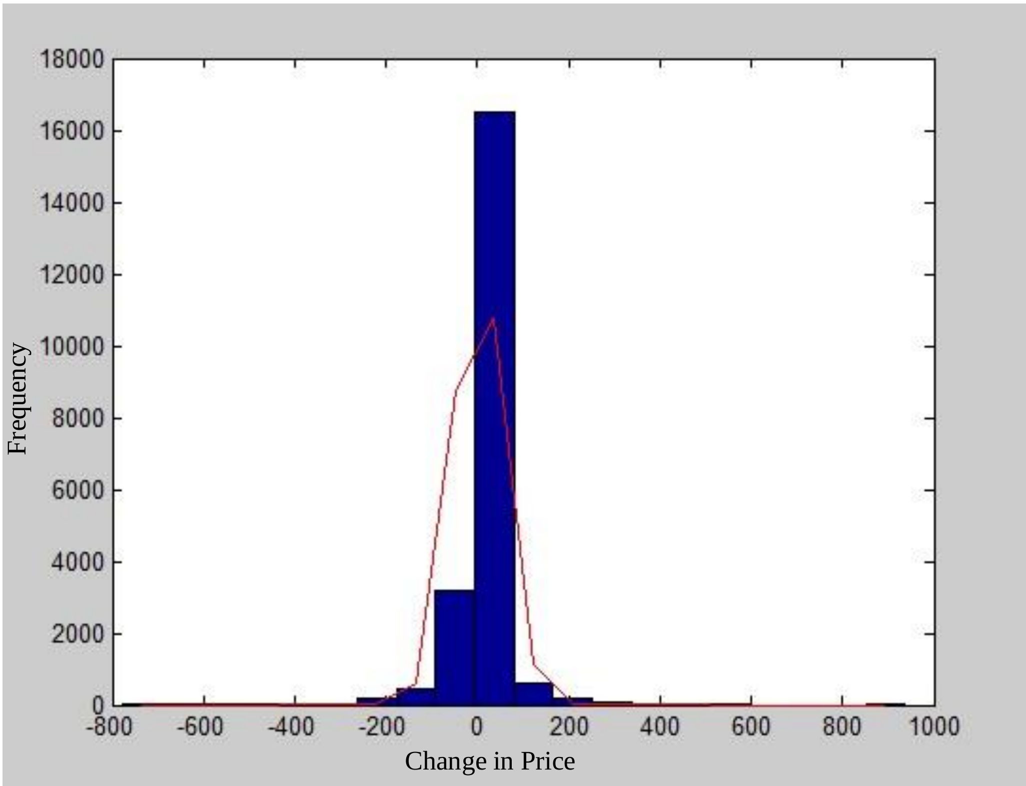

Figure 6: Histogram of Dow Price Change (All-Time)

We see that the average price change per day is 0.7138 with a standard deviation of 54.5409. The output plot shown below is not quite as normal as I had expected, but it does follow the general pattern with most days having little price change. Occasionally the Dow price changes drastically, but the frequency is low enough that the distribution is nearly unaffected.

The bars with highest frequency surround a zero-dollar change in price, which means that it is not uncommon for the stock market index to remain unchanged between days. However, notice that the mean change in price is positive. **We can expect the market to grow by about 0.71 per day in the long run**, which is good news for anyone who is wary of investing in stocks. With that said, the market is still somewhat volatile, and since we are analyzing many data points spanning many decades, our findings do not translate to short or shorter-term profits. Another guideline to keep in mind for the informed investor is that the stock market's future is not necessarily influenced by its historical highs and lows.

<u>Results for 2000-2013:</u>

Note that we tested the Dow on its entire lifespan, from 1928 to 2013. Many things have changed in between, so what if we ran the distribution test on only the last decade or so? Let's see the results of the Dow from 2000 until June 4, 2013.

```
xbar = 1.0905
s = 125.3886
```

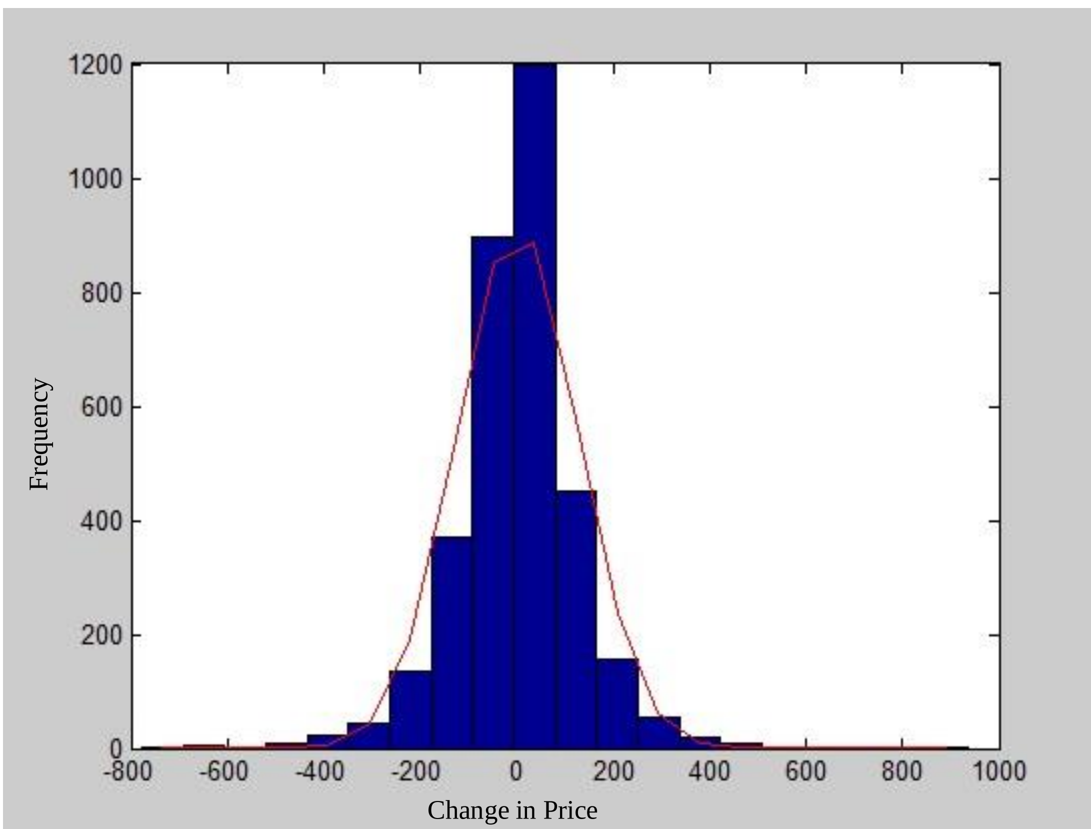

Figure 7: Histogram of Dow Price Change (Recent)

Interestingly, we find that the distribution of price change is now much more normal, at least according to our Gaussian fit. The **mean price change has increased to around 1.0905** as well, although the standard deviation increased with it. It seems that the risks have gotten slightly higher; the Dow both increases and decreases by more than before. But the histogram shows that every bar on the positive side of price change (that is, growth in the market) is slightly higher in frequency than its negative counterpart (down market). We can see the effects of a growing market in the current valuation of the Dow, as seen in Figure 8 below. It's no wonder that the market has grown so much in the past decade or so!

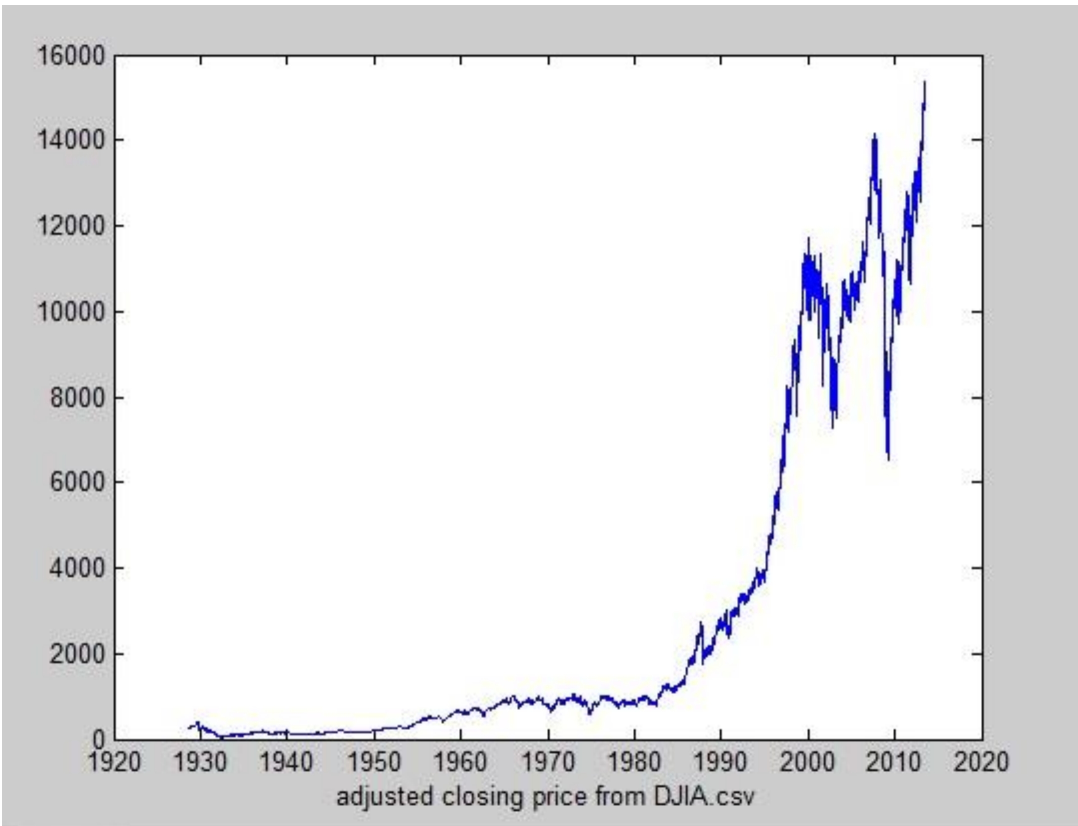

Figure 8: Dow Jones Stock Prices (1928-2013) in time-domain

TECHNICAL ANALYSIS

Some people have noticed that certain stocks behave in a rather cyclical manner. In other words, the changes in stock price often repeat the same trend over a period of time and it may even look like a sine wave. This is where **Fast Fourier Transform** (FFT) comes in. Instead of seeing the stock price in *time-domain* – that is, the stock price over time – we can skip over the overused charts and examine a company's shares in *frequency-domain.*

Many experts have tried using FFT to examine and even predict the stock market, but the outcome has been rather unclear. Using some of our code from the previous homework, we will demonstrate the effects of running Fast Fourier Transform on the Dow Jones Industrial Average.

Matlab Code:

```
function StockFFT(fileName)

%copy down the information into 'fileName.csv'
s=urlread(strcat('http://ichart.finance.yahoo.com/table.csv?s=',fileName));
fid=fopen(strcat(fileName,'.csv'),'wt');
fprintf(fid,s);

[time, quotes] = read_stock(strcat(fileName,'.csv'));
n = length(quotes);
power_spectrum = abs(fft(quotes)).^2;
frequencies = linspace(0, 1.0, n);

%% sampling frequency in days
plot(frequencies(2:floor(n/2)), power_spectrum(2:floor(n/2)));
xlabel('Frequency')
ylabel('Power')

%% sampling frequency in years
freqs = linspace(0, 252, length(power_spectrum));
xlabel('Frequency')
ylabel('Power')

figure; plot( freqs, power_spectrum );
```

Results:

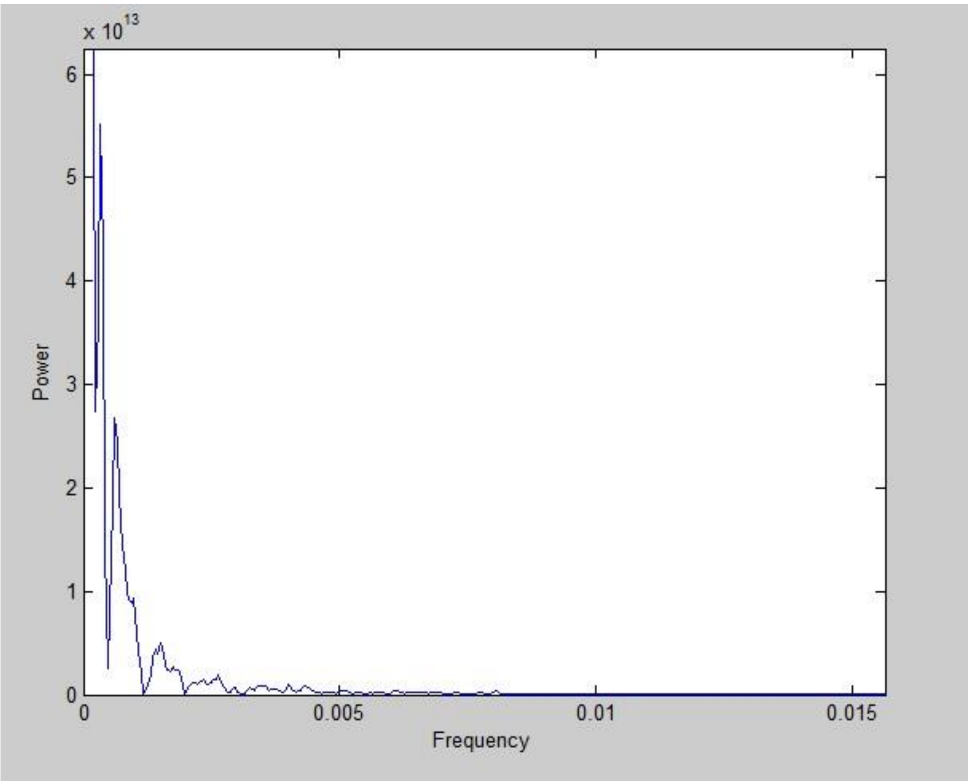

Figure 9: Power Spectrum of the Dow (FFT)

The graph above in Figure 9 is generated by squaring the absolute value of the Fast Fourier Transform, then mapping it against the "frequency" of the stock compressed into a single day. Fourier Transform expects us to provide a periodic signal, so if our resulting graph does not look ideal, then perhaps the stock data is not very periodic! However, we can see that the peaks coincide with cycles in our stock prices. Having a huge peak at a very small frequency makes sense, as changes in the Dow's stock price essentially complete tiny "cycles" in very short time periods due to what some call the "random nature" or perhaps even the "self-correction" of the stock market.

Knowing the period of time it takes for a specific company's stock to complete a cycle would theoretically allow investors to pinpoint the best times to buy and sell stock: buy shares at its lowest point and sell shares as it nears its top. But alas, stock prices are far from following a sinusoidal wave so people cannot know exactly what part of the market cycle we are currently in.

Even so, knowing how the stock market cycle works doesn't mean we that we should neglect the trend of stock prices as well. Using Matlab's tools like `polyfit` and `polyval`, we can generate a **polynomial regression** that acts much like a simple moving average. We utilize the `stock_read.m` code and add the best-fit regression line over the time-domain price graph.

Matlab Code:

```
function  [ time adjusted_close ] = StockRegression( CSV_Filename )

fid = fopen( CSV_Filename );
header = fgetl( fid );
data = textscan(fid,'%s%f%f%f%f%f%f','Delimiter',',','EmptyValue',-Inf);
fclose(fid);
s = data{1};
closing_price = data{7};
n = length(s);
time = zeros(n,1);
adjusted_close = zeros(n,1);
month_days = [ 0 31 28 31 30 31 30 31 31 30 31 30 ];
cumulative_month_days = cumsum( month_days );
for i=1:n
   Date = sscanf( s{n+1-i}, '%f-%f-%f' );
   time(i) = (Date(3) + cumulative_month_days(floor(Date(2))) + 365 *
Date(1)) / 365;
   adjusted_close(i) = closing_price(n+1-i);
end
figure
plot(time, adjusted_close)
xlabel(sprintf('adjusted closing price from %s',CSV_Filename))

%generate a polynomial regression that acts as an average
degree = 3;
xValue = linspace(time(1),time(length(time)));
coefficient = polyfit(time,adjusted_close,degree);
yValue = polyval(coefficient, xValue);
hold on
plot(xValue, yValue, 'r');
```

Results:

Most people think of linear regression lines of degree one when asked to find the best-fit of a graph. We hypothesize that a straight line most likely is not a good average fit for the Dow's price graph. Nevertheless, we will show what the linear regression line looks like.

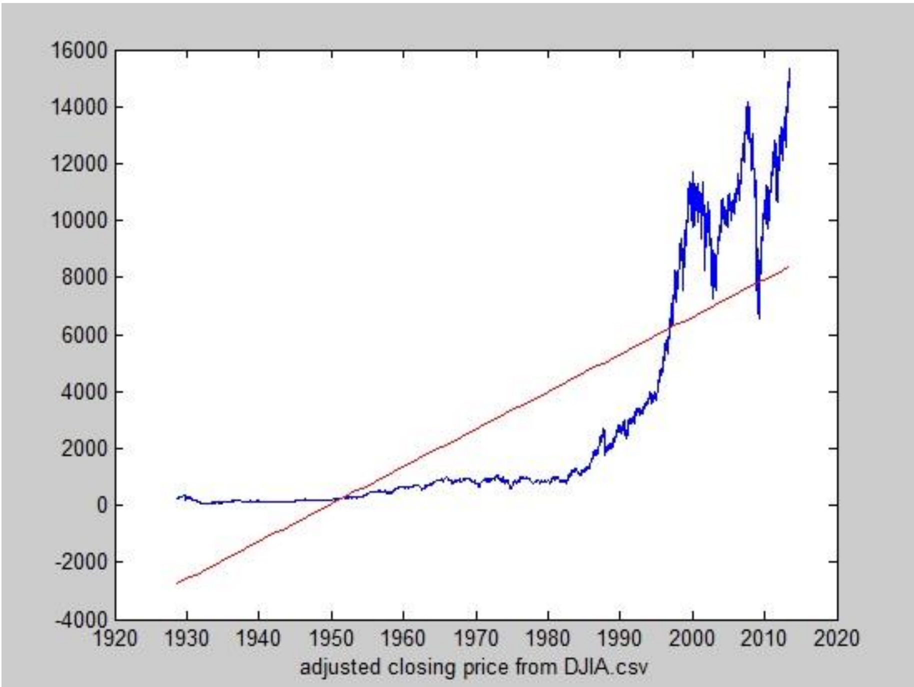

Figure 10: Polynomial Regression on the Dow (linear with degree 1)

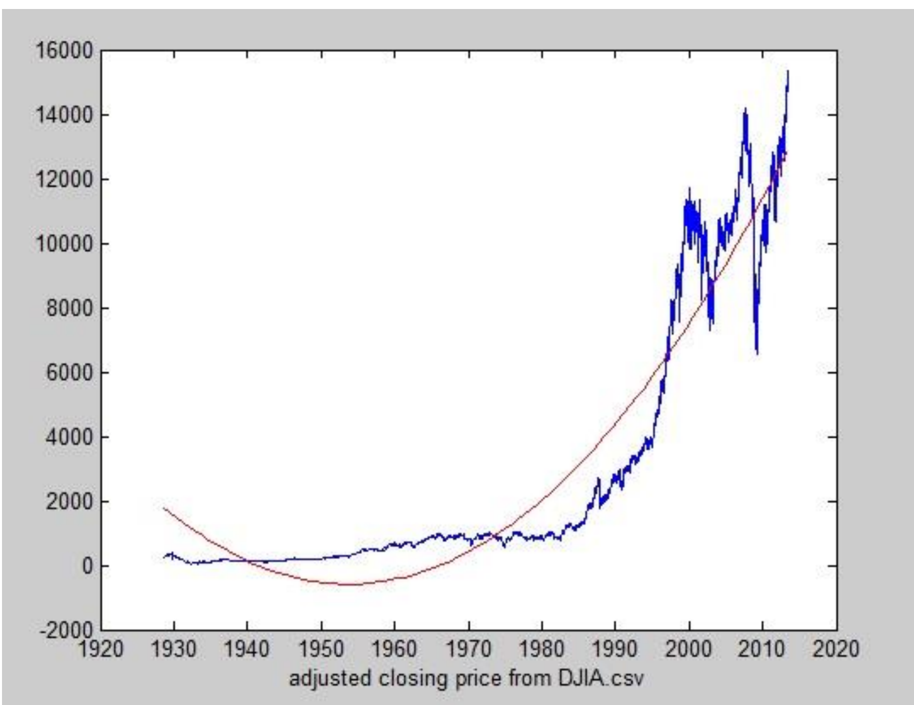

Figure 11: Polynomial Regression on the Dow (degree 2)

The regression line with degree 2 seems to be getting closer to the actual moving average of the share price. But we can tell that it is overshooting in the 1930s and 1980s-1990s, then underestimating the price between 1950 and 1970. It is not too far off, but the regression is less accurate since it has to incorporate the drastic growth of the 2000s stock market. Perhaps adding another degree will give us an even closer fit.

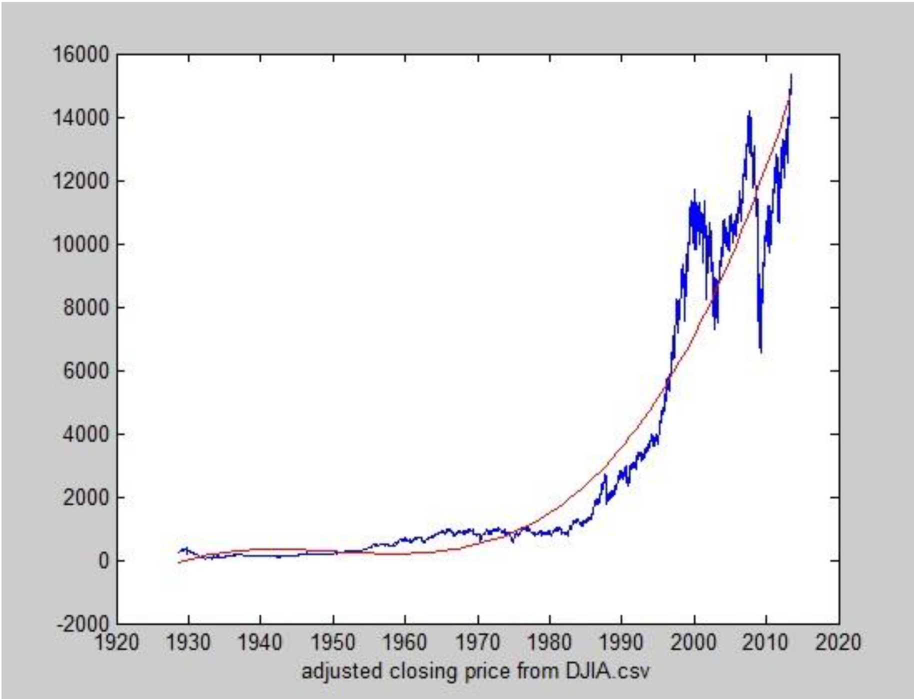

Figure 12: Polynomial Regression on the Dow (degree 3)

Now this graph looks much more accurate! If we increase the degree of the polynomial regression even more, the fit does not appear to be better, so a third-degree will suffice. As mentioned before, this polynomial regression seems just like a *simple moving average* (SMA).

Investors experienced in technical analysis of stocks often use the SMA to determine whether to buy or sell. If the current stock price has just dipped slightly below the SMA, it may be a good time to buy (provided the market is not expected to crash anytime soon). This is because our SMA tells us that it is likely the stock price will rebound to a higher price and create a profit. On the contrary, if a stock price is consistently under the SMA and it suddenly tips over the average, it may very well be an indication to sell: it will likely rebound and dip even more than before, and if we sell it at slightly above the SMA, the investor would be getting the most out of his money and situation.

Another technical indicator is *moving average convergence divergence* (MACD), which depends on the convergence and divergence of several different averages, as the name suggests. MACD has more potential to predict the future of the stock market, unlike other moving averages which are known as “lagging indicators” – useful for pointing out market trends after the fact. MACD includes a signal line and histogram along with its moving average. It has been noted that at the moment the histogram flips upright, it is a beneficial time to buy the stocks. Momentum usually carries the price even higher, and the investor can sell for a profit. The opposite is true as well; once the histogram flips downward, the stock price is in for a decrease, and selling now could mean less of a loss. Let's see this in terms of the Dow using a chart from Yahoo! Finance.

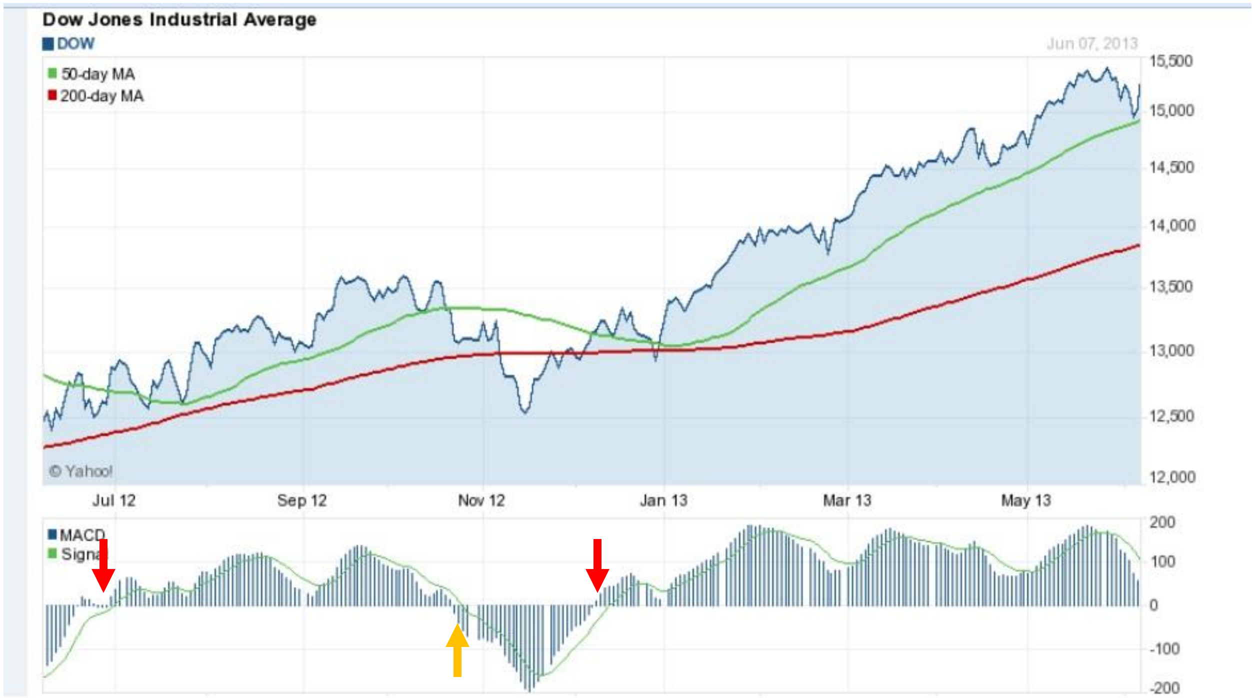

Figure 13: SMA and MACD of the Dow (Yahoo! Finance)

The arrows indicate ideal moments to buy or sell shares in the Dow. The red arrows are signals to buy, and as we can see, the actual share price increases from the rising momentum. The orange arrow is a signal to sell, as we note that the actual share price drops below the 50- and 200-day moving averages. Thus we see that technical analysis through Fast Fourier Transform and polynomial regressions can help find cyclical patterns in stock prices and determine whether to buy or sell a particular stock.

DIVERSIFYING YOUR PORTFOLIO

Now that we know the basics of the stock market, let us move on to selecting stocks to invest in. The saying goes: "don't put all your eggs in one basket." This rule applies especially in stock investing. A portfolio – a collection of stock investments – that is not at least slightly diverse in terms of industry or change in price has a small chance of making it big, but a large chance of losing lots of money. Having a well-balanced portfolio cuts risks while still retaining the opportunity to make gains.

But how can Matlab help us in this endeavor? Enter the **Coefficient Correlation**, a measure that ranges from -1 to 1. The former means two stocks are negatively correlated and a price hike in one stock means a price drop in the other. The latter means two stocks are positively correlated and it is not uncommon for both stocks to increase in price together. On the other hand, a coefficient of 0 means there is no known relation between the stocks.

Matlab Code:

```
function C = StockCorrelation

Stock1 = […];
Stock2 = […];

%Find correlation coefficient matrix
cMatrix = corrcoef(Stock1,Stock2);

%Pull out the coefficient constant
C = cMatrix(2);
```

Results:

We will find the correlation between Google (GOOG) and Apple (AAPL) from August 19, 2004 until June 3, 2013. The `Stock1` and `Stock2` matrices are essentially same-length vectors of stock prices which we manually extracted from the .csv files. In our sample test, we obtained the coefficient constant `C = 0.7730`, indicating that Google and Apple are at least somewhat similar in their returns. A profit in Google is usually accompanied with a profit in Apple, but the same phenomenon occurs for a net loss as well.

What if we tried calculating the coefficient correlation between companies of different industries? We'll keep the same timeframe for simplicity.

McDonald's and Apple:
```
C = 0.9354
```
Alcatel-Lucent and Apple:
```
C = -0.7211
```
JP Morgan and Coca-Cola:
```
C = 0.5040
```
Bank of America and Boeing:
```
C = 0.1834
```

As we can see, McDonald's and Apple both have similar patterns of price change. In contrast, when Apple stock grows, Alcatel-Lucent (a French telecommunications company and owner of Bell) is rather likely to fall. It seems that most Dow companies have a positive correlation with each other, perhaps because all of the companies are large in size and their stock is more stable.

Diversification of a portfolio should not only be limited to correlation between past returns, especially since the future of the stock market does not necessarily follow the past. Stocks should also be picked based on the company's market cap, industry, and even its location. The extent to which your portfolio is diversified is up to personal preference; the correlation coefficient between chosen stocks is simply an indicator of how diversified your portfolio is. It is beneficial for your portfolio to have a spread of various different coefficients, but each individual stock should still be a solid pick based on the company's financials. If your portfolio is diversified but made up of poor choices, you will still lose when the companies go bankrupt.

CONCLUSION

So there we have the fundamentals of the stock market summed up in terms of Matlab and technical analysis. We have only scratched the surface of investment strategies, but that is barely the point. The big deal is that we have successfully used Matlab's matrix tools to demonstrate the market in practice. We first saw how the market is measured in terms of growth. To put it in investor terms: is it a *bull market* in which shares are increasing in value? Or are we facing a *bear market* where most stocks are undergoing devaluation? Indices like the Dow are the answer here, as they generalize the stock market and give investors a window that summarizes price change. Can indices give us a false perspective of the market as a whole? Absolutely. But it is highly unlikely that several indices would disagree with the actual market, so we depend on the Dow to measure the market change. Principal component analysis tells us that the Dow comprises large-volume companies with medium-to-high share prices, so these giant corporations should represent a fair share of the market. Generating a histogram of Dow price changes provides evidence that the stock market is not completely random. While the distribution is relatively normal, we see that the mean is actually positive, so the market is indeed growing over time.

Next we used some technical analysis methods on the Dow. Plotting the Fast Fourier Transform showed us that the stock market price over time does form some cycles, though perhaps they are small and not very reliable. Nevertheless, this lends credibility to the idea that the stock market can be "predicted" in a sense using technical analysis. The polynomial regression allows us to better grasp how the stock has changed over history and maps that onto the time-domain of stock prices. Using the regression as a guideline, investors will have a better idea of when to buy or sell specific stocks for maximum profit.

Lastly, we attempted to create a portfolio of diverse stocks using two matrices' correlation coefficient. It is unwise to pick stocks to invest in purely because of the correlation coefficient. The company of interest should first and foremost have trustworthy financials and a positive forward outlook. But the coefficient should be taken into consideration prior to the purchase of stocks. That way, the investor can be assured that he is not gambling away money by purchasing stocks in three different companies that have almost the same correlated returns.

Despite having all these technical tools at hand, the stock market is still not to be taken lightly. All stocks are held at the mercy of volatility, and it is doubtful that our technology can precisely predict the next market crash or even the next bear market – especially if it is based on irrational human behavior. Pick your stocks wisely, be prepared to hold on to them for some time (as the market grows in the long-term), and remember to use these technical resources to supplement your decision-making in stock market investing!

BIBLIOGRAPHY

“DJI: Summary for Dow Jones Industrial Average.” *Yahoo! Finance*. Yahoo!. Web. 6 June 2013.

Kelly, Jason. *The Neatest Little Guide to Stock Market Investing.* 2013 ed. New York: PLUME, 2012. Print.

Parker, D.S. *Exporing the Matrix – Adventures in Modeling with MATLAB*. Los Angeles: UCLA Course Reader Solutions, 2013. Print.