Evaluation Bench for Portfolio Optimization

A Major Qualifying Project Report

submitted to the Faculty of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

by

_____

Emilia Daniela Holban

_____

_____

Professor Domokos Vermes, Advisor

Abstract

This project implements a MATLAB based evaluation workbench to assess and compare portfolio management strategies. Daily data for 170 stocks carefully selected to cover the market between Dec 24 1998 and March 31, 2008 forms the basis of the evaluation. Quarterly or yearly rebalancing is implemented. Three strategies are compared in the current implementation – uniform, top quartile and optimized, minimum variance portfolios.

Acknowledgments

The outmost appreciation goes towards Professor Domokos Vermes for his constant support, advice and valuable guidance.

Table of Contents

Introduction

This project was intended to be a test bench for portfolio optimization – a system to which new strategies can be added such that their efficiency of maximizing value can be measured against other strategies.

The portfolio used in this project was chosen such that it was homogenous and large enough to represent the market and would have enough data history to enable analysis of the data over a period of nine years. The common market indeces, such as the Dow Jones Industrial Average or the S&P500 were not a good choice since in the past nine years they contained companies which went bankrupt during the bust of the internet bubble, other which got bought out or recent IPO's, like Google. These companies do not have long enough historical data to be used on a study that tests trading strategies implemented over the last nine years, since systematic investment in them is impossible.

The trading baseline strategies implemented were to uniformly distribute money over the entire portfolio at the end of each year or quarter or to invest in the top 40 companies of the past year or quarter. Other strategies would be compared to the results provided by these two initial strategies. In order to exemplify this, a third strategy was implemented -   portfolio optimization to minimize risk subject to a minimum required return which was compared to the previous two.

1. Portfolio Selection

Goals

In order to attempt finding a diversified portfolio that would clearly exhibit the results of different optimization methods, a strategy needed to be devised to help with the choice of stocks composing the testing portfolio. The goal was to choose a portfolio that would satisfy the properties:

- large enough

- long enough history

- homogenous portfolio

The size of the portfolio was chosen such that it is **large enough** to represent the entire industry, but small enough to make repeated analysis on the data feasible. For example, if the portfolio included 50 stocks, then there would have existed the risk that they were not representing the industry well enough and then the portfolio would have not been uniformly distributed over the industry. In this case results would be skewed because parts of the market and their impact on a portfolio would be ignored. A very large portfolio, for example, one containing 2000 stocks could cover all areas of the industry, but would make computational analysis on the data associated with it take a long time every run, hindering research. This is why a portfolio of 170 stocks was chosen, picked such that it satisfies all the goals established.

The trading period chosen was from the beginning of 1999 to the end of the first quarter in 2008. For the data to be analyzed and in order to avoid large amounts of missing data giving inconclusive results, the companies in the portfolio were chosen such

that they had **long enough historical data,** their first trading day needing to be before the end of 1998.

The portfolio was chosen in such a way that it is **homogenous** over market capitalization (its total stock market value), style (value, growth and blend) and market sectors such that the results would not be affected by investment in only one or a handful of areas of the market. This way the portfolio is affected only by market forces, and not by events that affected only small areas of the market.

The Market Universe for the Present Study

The choice of stocks was made from the 2000 monitored by Morningstar, Inc. (www.morningstar.com). Morningstar is an investment research company whose financial data is often used by financial advisors and analysts. The way this company divides the stocks traded in the New York Stock Exchange is by placing each in one of nine categories, based on Value, Core and Growth, versus Small, Medium or Large market capitalization. Using this classification, the stocks were divided into the following nine categories:

Figure 1. Morningstar division table for the NYSE traded companies

These 2000 stocks constitute the "universe" out of which a data sample of stocks would be chosen to form the portfolio.

Restricting the portfolio to contain stocks which have enough historical data (going back to December 24'th, 1998) was a necessary condition. The reason this date was chosen has to do with the trading strategy. The year 1999 was considered a "buffer year", a period in which enough data is gathered in order to start making more informed investment decisions. This data is passed to a Matlab program written for this project, called the *Trader*. The trading period lasts 5 days, which means that the sum invested in the portfolio would be spread out evenly into five days such that to avoid a lucky or unlucky trading day that would affect the data. In order to start the new quarter/annual time periods from Jan 1st 1999, the last five trading days of 1998 were also included in the data. Having a long enough history assures that the analysis performed on the stocks is conclusive enough to affect trading. The downside to this is that it creates a bias – companies with new public offerings, bought out or bankrupt are not contained.

Stocks are also divided by market sectors, falling in one of the following categories: Basic Materials, Conglomerates, Consumer Goods, Financial, Healthcare, Industrial Goods, Services, Technology and Utilities. In each market sector there are subdivisions called industries.

Here is the breakdown of the market sectors:

Basic Materials

Agricultural Chemicals
Aluminum
Chemicals - Major Diversified
Copper
Gold
Independent Oil & Gas
Industrial Metals & Minerals
Major Integrated Oil & Gas
Nonmetallic Mineral Mining
Oil & Gas Drilling & Exploration
Oil & Gas Equipment & Services
Oil & Gas Pipelines
Oil & Gas Refining & Marketing
Silver
Specialty Chemicals
Steel & Iron
Synthetics

Conglomerates

Conglomerates

Consumer Goods

Appliances
Auto Manufacturers - Major
Auto Parts
Beverages - Brewers
Beverages - Soft Drinks
Beverages - Wineries & Distillers
Business Equipment
Cigarettes
Cleaning Products
Confectioners
Dairy Products
Electronic Equipment

Farm Products
Food - Major Diversified
Home Furnishings & Fixtures
Housewares & Accessories
Meat Products
Office Supplies
Packaging & Containers
Paper & Paper Products
Personal Products
Photographic Equipment & Supplies
Processed & Packaged Goods
Recreational Goods, Other
Recreational Vehicles
Rubber & Plastics
Sporting Goods
Textile - Apparel Clothing
Textile - Apparel Footwear &

Accessories

Tobacco Products, Other
Toys & Games
Trucks & Other Vehicles

Financial
Accident & Health Insurance
Asset Management
Closed-End Fund - Debt
Closed-End Fund - Equity
Closed-End Fund - Foreign
Credit Services
Diversified Investments
Foreign Money Center Banks
Foreign Regional Banks
Insurance Brokers
Investment Brokerage - National
Investment Brokerage - Regional
Life Insurance
Money Center Banks
Mortgage Investment
Property & Casualty Insurance
Property Management
REIT - Diversified
REIT - Healthcare Facilities
REIT - Hotel/Motel
REIT - Industrial
REIT - Office
REIT - Residential
REIT - Retail
Real Estate Development
Regional - Mid-Atlantic Banks
Regional - Midwest Banks
Regional - Northeast Banks
Regional - Pacific Banks
Regional - Southeast Banks
Regional - Southwest Banks
Savings & Loans
Surety & Title Insurance

Healthcare

Biotechnology
Diagnostic Substances
Drug Delivery
Drug Manufacturers - Major
Drug Manufacturers - Other
Drug Related Products
Drugs - Generic
Health Care Plans
Home Health Care

Hospitals
Long-Term Care Facilities
Medical Appliances & Equipment
Medical Instruments & Supplies
Medical Laboratories & Research
Medical Practitioners
Specialized Health Services

Industrial Goods

Aerospace/Defense - Major Diversified
Aerospace/Defense Products & Services
Cement
Diversified Machinery
Farm & Construction Machinery
General Building Materials
General Contractors
Heavy Construction
Industrial Electrical Equipment
Industrial Equipment & Components
Lumber, Wood Production
Machine Tools & Accessories
Manufactured Housing
Metal Fabrication
Pollution & Treatment Controls
Residential Construction
Small Tools & Accessories
Textile Industrial
Waste Management

Services

Advertising Agencies
Air Delivery & Freight Services
Air Services, Other
Apparel Stores
Auto Dealerships
Auto Parts Stores
Auto Parts Wholesale
Basic Materials Wholesale
Broadcasting - Radio
Broadcasting - TV
Building Materials Wholesale
Business Services
CATV Systems
Catalog & Mail Order Houses
Computers Wholesale
Consumer Services
Department Stores
Discount, Variety Stores
Drug Stores
Drugs Wholesale
Education & Training Services
Electronics Stores
Electronics Wholesale
Entertainment - Diversified
Food Wholesale
Gaming Activities
General Entertainment
Grocery Stores
Home Furnishing Stores
Home Improvement Stores
Industrial Equipment Wholesale
Jewelry Stores
Lodging
Major Airlines
Management Services
Marketing Services
Medical Equipment Wholesale
Movie Production, Theaters

Music & Video Stores
Personal Services
Publishing - Books
Publishing - Newspapers
Publishing - Periodicals
Railroads
Regional Airlines
Rental & Leasing Services
Research Services
Resorts & Casinos
Restaurants
Security & Protection Services
Shipping
Specialty Eateries
Specialty Retail, Other
Sporting Activities
Sporting Goods Stores
Staffing & Outsourcing Services
Technical Services
Toy & Hobby Stores
Trucking
Wholesale, Other

Diversified Computer Systems
Diversified Electronics
Healthcare Information Services
Information & Delivery Services
Information Technology Services
Internet Information Providers
Internet Service Providers
Internet Software & Services
Long Distance Carriers
Multimedia & Graphics Software
Networking & Communication Devices
Personal Computers
Printed Circuit Boards
Processing Systems & Products
Scientific & Technical Instruments
Security Software & Services
Semiconductor - Broad Line
Semiconductor - Integrated Circuits
Semiconductor - Specialized
Semiconductor Equipment & Materials
Semiconductor- Memory Chips
Technical & System Software
Telecom Services - Domestic
Telecom Services - Foreign
Wireless Communications

Technology

Application Software
Business Software & Services
Communication Equipment
Computer Based Systems
Computer Peripherals
Data Storage Devices
Diversified Communication Services

Utilities

Diversified Utilities
Electric Utilities
Foreign Utilities
Gas Utilities
Water Utilities

Subdivision into market sectors and the nine Morningstar categories provides two different ways to categorize stocks across the market. The market sector category divides the stocks according to the activity of the company, while the Morningstar style box presented earlier divides the 2000 stocks of the "universe" into nine categories, based on the market capitalization of the company and its style. Combining these two methods of classifying stocks, one can get a uniform portfolio across the market.

Provided with the list of stocks in each sector and industry and also with the list of stocks in each of the Morningstar categories, there was need for the information to be cross-referenced, along with information about the first trading day of the company. Having this data would facilitate uniformly picking out of each of the nine Morningstar

categories around three stocks per Sector pertaining of different Industries. The chosen stocks also needed to have enough history.

Data gathering

By using a program written in Python for this project, the stocks in each of the nine categories developed by Morningstar have been cross-referenced with the industry and sector information, also adding for each the first trading date. What resulted were nine tables, each containing the stocks from each Morningstar category. Each stock had attached the industry and sector it was part of and its first trading day. The reason why the trading day was added was so that stocks without enough trading history not to be considered towards the portfolio. The first trading day was determined by using another program written for this program that would retrieve all trading history for a given set of stocks.

An average of three stocks was chosen from each industry per table, with the condition that they had historical trading data going back to the end of 1998. This resulted in a portfolio of 170 stocks chosen uniformly over style, market capital, market sectors and with enough historical data.

Once the portfolio was chosen, the tickers were put into a file. A program written in Python for this project created a .csv file with all the adjusted daily closing price of each stock starting from December 24'th, 1998. The data was taken from the finance.yahoo.com site.

This is the content of the portfolio, along with the Sector and first trading day:

11

Symbol	Name	Sector	Industry	Morningstar Category
ADBE	Adobe Systems Inc.	Application Software	Technology	MLGRT
AES	AES Corporation	Electric Utilities	Utilities	MLGRT
AET	Aetna Inc.	Health Care Plans	Healthcare	MLCRT
ATG	AGL Resources Inc.	Gas Utilities	Utilities	MMVLT
AA	Alcoa Inc.	Aluminum	Basic Materials	MLVLT
MO	Altria Group Inc.	Cigarettes	Consumer Goods	MLVLT
ABC	AmerisourceBergen Corporation	Drugs Wholesale	Services	MMCRT
APA	Apache Corporation	Independent Oil & Gas	Basic Materials	MLVLT
AMAT	Applied Materials	Semiconductor Equip/Matrl	Technology	MLGRT
ACI	Arch Coal Inc.	Industrial Metals/Mineral	Basic Materials	MMGRT
ADSK	Autodesk Inc.	Technical/System Software	Technology	MMGRT
AN	AutoNation Inc.	Auto Dealerships	Services	MMVLT
BRL	Barr Pharmaceuticals Inc.	Drugs - Generic	Healthcare	MMCRT
BAX	Baxter International Inc.	Medical Instruments/Supls	Healthcare	MLGRT
BIO	Bio-Rad Laboratories Inc.	Scientific/Tech Instrmnts	Technology	MSCRT
BBOX	Black Box Corporation	Networking & Commun Dvcs	Technology	MSVLT
BWA	BorgWarner Inc.	Auto Parts	Consumer Goods	MMVLT
BYD	Boyd Gaming Corporation	Resorts & Casinos	Services	MMCRT
BRC	Brady Corporation	Business Services	Services	MSCRT
BGG	Briggs & Stratton Corporation	Diversified Machinery	Industrial Goods	MSVLT
BRS	Bristow Group Inc.	Air Services Other	Services	MSVLT
BRKL	Brookline Bancorp Inc.	Savings & Loans	Financial	MSCRT
CBT	Cabot Corporation	Specialty Chemicals	Basic Materials	MMCRT
CDNS	Cadence Design Systems	Technical/System Software	Technology	MMCRT
CWT	California Water Service Group	Water Utilities	Utilities	MSVLT
CBM	Cambrex Corporation	Biotechnology	Healthcare	MSCRT
KMX	CarMax Inc.	Auto Dealerships	Services	MMGRT
CAE	Cascade Corporation	Farm/Construction Machnry	Industrial Goods	MSCRT
CEDC	Central European Distribution Corporation	Beverages-Winery/Distlers	Consumer Goods	MSGRT
CHRW	CH Robinson Worldwide Inc.	Air Delivery/Freight Svcs	Services	MMGRT
CHK	Chesapeake Energy Corp.	Independent Oil & Gas	Basic Materials	MLCRT
CBK	Christopher & Banks Corporation	Apparel Stores	Services	MSCRT
C	Citigroup Inc.	Money Center Banks	Financial	MLVLT
CZN	Citizens Communications Company	Telecom Services/Domestic	Technology	MMVLT
CLHB	Clean Harbors Inc.	Waste Management	Industrial Goods	MSGRT
CLX	Clorox Company	Cleaning Products	Consumer Goods	MMCRT
KO	Coca-Cola Company	Beverages-Soft Drinks	Consumer Goods	MLCRT
CL	Colgate-Palmolive Company	Personal Products	Consumer Goods	MLGRT
ED	Consolidated Edison Company	Diversified Utilities	Utilities	MLVLT
CEG	Constellation Energy Group Inc.	Electric Utilities	Utilities	MLCRT
CMI	Cummins Inc.	Diversified Machinery	Industrial Goods	MMCRT

DECK	Deckers Outdoor Corporation	Textile-Apparel Footwr/Ac	Consumer Goods	MSGRT
DE	Deere & Company	Farm/Construction Machnry	Industrial Goods	MLCRT
XRAY	DENTSPLY International Inc.	Medical Instruments/Supls	Healthcare	MMCRT
DO	Diamond Offshore Drilling Inc.	Oil & Gas Drilling/Explor	Basic Materials	MMGRT
DST	DST Systems Inc.	Information/Delivery Svcs	Technology	MMCRT
DD	DuPont EI de Nemours & Co	Agricultural Chemicals	Basic Materials	MLCRT
DY	Dycom Industries Inc.	Heavy Construction	Industrial Goods	MSCRT
SSP	E.W. Scripps Company	Publishing - Newspapers	Services	MMCRT
EXP	Eagle Materials Inc.	Cement	Industrial Goods	MSGRT
EMN	Eastman Chemical Company	Chemicals-Major Diversifd	Basic Materials	MMVLT
EK	Eastman Kodak Company	Photographic Equip/Supls	Consumer Goods	MMVLT
EV	Eaton Vance Corporation	Asset Management	Financial	MMGRT
EE	El Paso Electric Company	Electric Utilities	Utilities	MSCRT
EDS	Electronic Data Systems Corporation	Information Technlgy Svcs	Technology	MLGRT
EQT	Equitable Resources Inc.	Gas Utilities	Utilities	MMCRT
ELS	Equity Lifestyle Properties Inc.	REIT - Residential	Financial	MSGRT
EQR	Equity Residential	REIT - Residential	Financial	MLGRT
EL	Estee Lauder Companies	Personal Products	Consumer Goods	MMCRT
FAST	Fastenal Company	General Building Matrials	Industrial Goods	MMGRT
FDX	FedEx Corporation	Air Delivery/Freight Svcs	Services	MLGRT
FBP	First BanCorp	Foreign Regional Banks	Financial	MSCRT
F	Ford Motor Company	Auto Manufacturers/Major	Consumer Goods	MLVLT
FST	Forest Oil Corporation	Independent Oil & Gas	Basic Materials	MMVLT
FELE	Franklin Electric Co.	Industrial Electrical Eqp	Industrial Goods	MSGRT
BEN	Franklin Resources	Asset Management	Financial	MLGRT
GCI	Gannett Co. Inc.	Publishing - Newspapers	Services	MLVLT
GPS	Gap Inc.	Apparel Stores	Services	MLVLT
DNA	Genentech Inc.	Biotechnology	Healthcare	MLGRT
GWR	Genesee & Wyoming Inc.	Railroads	Services	MSGRT
GLF	GulfMark Offshore Inc.	Oil & Gas Equipment/Svcs	Basic Materials	MSCRT
FUL	H.B. Fuller Company	Specialty Chemicals	Basic Materials	MSCRT
HANS	Hansen Natural Corporation	Beverages-Soft Drinks	Consumer Goods	MMGRT
HHS	Harte-Hanks Inc.	Marketing Services	Services	MSVLT
HIG	Hartford Financial Services Group Inc.	Prprty/Casualty Insurance	Financial	MLVLT
HLS	Healthsouth Corporation	Specialized Health Svcs	Healthcare	MSVLT
MLHR	Herman Miller Inc.	Business Equipment	Consumer Goods	MMGRT
HES	Hess Corporation	Oil & Gas Refining/Mrktng	Basic Materials	MLVLT
HPQ	Hewlett-Packard Company	Diversified Computer Sys	Technology	MLCRT
HOV	Hovnanian Enterprises Inc.	Residential Construction	Industrial Goods	MSVLT
IHP	IHOP Corp.	Restaurants	Services	MSCRT
ITW	Illinois Tool Works Inc.	Diversified Machinery	Industrial Goods	MLCRT
IR	Ingersoll-Rand Company Ltd.	Diversified Machinery	Industrial Goods	MLVLT
INTC	Intel Corporation	Semiconductor-Broad Line	Technology	MLCRT
ISCA	International Speedway Corporation	Sporting Activities	Services	MMVLT
IO	ION Geophysical Corporation	Scientific/Tech Instrmnts	Technology	MSGRT

JKHY	Jack Henry & Associates	Business Software & Svcs	Technology	MSCRT
JAH	Jarden Corporation	Housewares & Accessories	Consumer Goods	MMCRT
KG	King Pharmaceuticals Inc.	Drug Manufacturers/Other	Healthcare	MMVLT
LANC	Lancaster Colony Corporation	Food - Major Diversified	Consumer Goods	MSVLT
LEH	Lehman Brothers Holdings Inc.	Investmnt Brokerage-Natl	Financial	MLCRT
LXK	Lexmark International Inc.	Computer Peripherals	Technology	MMVLT
LECO	Lincoln Electric Holdings Inc.	Small Tools & Accessories	Industrial Goods	MMCRT
LNC	Lincoln National Corp.	Life Insurance	Financial	MLVLT
LMT	Lockheed Martin Corporation	Aerospace/Defense-Prd/Svc	Industrial Goods	MLCRT
LZ	Lubrizol Corporation	Specialty Chemicals	Basic Materials	MMVLT
MAC	Macerich Company	REIT - Retail	Financial	MMGRT
M	Macy's	Department Stores	Services	MLVLT
MAN	Manpower Inc.	Staffing/Outsourcing Svcs	Services	MMCRT
MKL	Markel Corporation	Prprty/Casualty Insurance	Financial	MMCRT
MAS	Masco Corporation	Lumber Wood Production	Industrial Goods	MMVLT
MTZ	MasTec Inc.	Heavy Construction	Industrial Goods	MSVLT
MRK	Merck & Co. Inc.	Drug Manufacturers/Major	Healthcare	MLCRT
MFA	MFA Mortgage Investments Inc.	Mortgage Investment	Financial	MSVLT
MGEE	MGE Energy Inc.	Diversified Utilities	Utilities	MSVLT
MCHP	Microchip Technology Inc.	Semiconductor-Specialized	Technology	MMGRT
MLNM	Millennium Pharmaceuticals Inc.	Drug Manufacturers/Other	Healthcare	MMGRT
MUR	Murphy Oil Corporation	Oil & Gas Refining/Mrktng	Basic Materials	MMCRT
MYGN	Myriad Genetics Inc.	Diagnostic Substances	Healthcare	MSGRT
NJR	New Jersey Resources Corporation	Gas Utilities	Utilities	MSVLT
NFX	Newfield Exploration Company	Independent Oil & Gas	Basic Materials	MMCRT
NEM	Newmont Mining	Gold	Basic Materials	MLGRT
NKE	Nike Inc.	Textile-Apparel Footwr/Ac	Consumer Goods	MLGRT
NUE	Nucor Corp.	Steel & Iron	Basic Materials	MLCRT
OGE	OGE Energy Corp	Electric Utilities	Utilities	MMVLT
ORI	Old Republic International	Surety/Title Insurance	Financial	MMVLT
PMTC	Parametric Technology Corporation	Technical/System Software	Technology	MSCRT
PDLI	PDL BioPharma Inc.	Biotechnology	Healthcare	MMGRT
PRGO	Perrigo Company	Drug Related Products	Healthcare	MSCRT
RL	Polo Ralph Lauren Corporation	Textile-Apparel Clothing	Consumer Goods	MMGRT
PX	Praxair Inc.	Synthetics	Basic Materials	MLGRT
PCP	Precision Castparts Corp.	Metals Fabrication	Industrial Goods	MLGRT
PGI	Premiere Global Services Inc.	Diversified Commun Svcs	Technology	MSVLT
PG	Procter & Gamble Company	Personal Products	Consumer Goods	MLCRT
RRC	Range Resources Corporation	Independent Oil & Gas	Basic Materials	MMGRT
RCRC	RC2 Corporation	Toys & Games	Consumer Goods	MSCRT
REG	Regency Centers Corporation	REIT - Retail	Financial	MMCRT
RS	Reliance Steel and Aluminum	Metals Fabrication	Industrial Goods	MMVLT
ROP	Roper Industries Inc.	Industrial Equip/Componts	Industrial Goods	MMGRT
SLE	Sara Lee Corporation	Processed/Packaged	Consumer Goods	MLVLT

		Goods		
SCG	Scana Corporation	Diversified Utilities	Utilities	MMVLT
SRP	Sierra Pacific Resources Holding Company	Diversified Utilities	Utilities	MMCRT
SLGN	Silgan Holdings Inc.	Packaging & Containers	Consumer Goods	MSVLT
SPG	Simon Property Group Inc.	REIT - Retail	Financial	MLCRT
SFD	Smithfield Foods Inc.	Meat Products	Consumer Goods	MMVLT
BID	Sotheby's Holdings Inc.	Business Services	Services	MMGRT
SEH	Spartech Corporation	Rubber & Plastics	Consumer Goods	MSCRT
S	Sprint Nextel Corporation	Wireless Communications	Technology	MLCRT
JOE	St. Joe Corporation	Real Estate Development	Financial	MMCRT
SBUX	Starbucks Corporation	Specialty Eateries	Services	MLGRT
HOT	Starwood Hotels & Resorts Worldwide Inc.	Lodging	Services	MLGRT
STT	State Street Corporation	Regional-Northeast Banks	Financial	MLCRT
SY	Sybase Inc.	Application Software	Technology	MMCRT
AMTD	TD Ameritrade Holding Corporation	Investmnt Brokerage-Natl	Financial	MMGRT
TEX	Terex Corporation	Farm/Construction Machnry	Industrial Goods	MMCRT
TWX	Time Warner Inc.	Entertainment - Diversified	Services	MLCRT
TJX	TJX Companies	Department Stores	Services	MLCRT
TG	Tredegar Corporation	Aluminum	Basic Materials	MSCRT
TGI	Triumph Group Inc.	Aerospace/Defense-Prd/Svc	Industrial Goods	MSCRT
TRST	TrustCo Bank Corporation of NY	Regional-Northeast Banks	Financial	MSVLT
TYC	Tyco International Ltd.	Diversified Electronics	Technology	MLVLT
UNP	Union Pacific Corporation	Railroads	Services	MLCRT
UTX	United Technologies	Conglomerates	Conglomerates	MLCRT
UNH	UnitedHealth Group Inc.	Health Care Plans	Healthcare	MLGRT
UHS	Universal Health Services	Hospitals	Healthcare	MMCRT
UNM	Unum Group	Accident/Health Insurance	Financial	MMVLT
USG	USG Corporation	General Building Matrials	Industrial Goods	MMVLT
VZ	Verizon Communications Inc.	Telecom Services/Domestic	Technology	MLVLT
VSAT	ViaSat Inc.	Communication Equipment	Technology	MSGRT
VPHM	ViroPharma Inc.	Biotechnology	Healthcare	MSGRT
VSH	Vishay Intertechnology	Diversified Electronics	Technology	MMVLT
BER	W.R. Berkley Corporation	Prprty/Casualty Insurance	Financial	MMVLT
WLT	Walter Industries Inc.	Industrial Metals/Mineral	Basic Materials	MSVLT
WMI	Waste Management Inc.	Waste Management	Industrial Goods	MLVLT
WY	Weyerhaeuser Company	Lumber Wood Production	Industrial Goods	MLVLT
WMB	Williams Companies Inc.	Oil & Gas Pipelines	Basic Materials	MLGRT
WTFC	Wintrust Financial Corporation	Regional-Midwest Banks	Financial	MSCRT
WWY	Wm. Wrigley Jr. Company	Confectioners	Consumer Goods	MLGRT
WYE	Wyeth	Drug Manufacturers/Major	Healthcare	MLCRT
XRX	Xerox Corporation	Business Equipment	Consumer Goods	MLCRT

2. Data Cleaning

When considering the history of stocks, some stocks had missing or wrong data for certain trading days. A method to fill in this missing data was required such that the missing or wrong data does not affect the trading or analysis.

To prevent errors in the data, rules about correct data were put in place – a spike over 20% in the price of a stock and then a return of over 20% in the span of 3 days was considered erroneous data and was deleted and replaced using the same method for missing data.

This method consists of replacing the missing data by a scaled new data that takes into account the historical variation. In order to keep the randomness of the data, a scalar value drawn from a normal distribution with mean 0 and standard deviation 1 is used.

Example: Imagine that for stock i, the data set looks like this:

Stock i

110.64	50.09	**15.27**	100.31	46.9
107.46	50.28	**15.1**	93.13	47.42
105.89	49.37	**15.04**	90.4	47.81
105.03	47.04	**15.17**	89.37	45.89
107.81	49.33	**15.19**	95.32	47.18
111.53	49.89	**15.76**	101.46	49.1
118.43	51.52	**15.84**	103.16	50.33	← Point2
113.56	50.42	**NaN**	101.08	50.53
108.29	49.34	**NaN**	97.62	49.56
118.71	50.02	**NaN**	99.99	50.74
124.15	49.99	**15.94**	106.24	49.09	← Point1
124.28	51.25	**15.75**	103.2	48.64
119.33	50.73	**15.59**	110.33	48.85
124.49	52.82	**15.89**	117.56	51.55
119.54	53.04	**15.76**	111.35	50.81
118.19	52.69	**15.65**	112.39	51.24
119.07	52.86	**15.99**	115.06	51.11
116.11	52.41	**15.65**	109.79	49.33

Figure 2. Data Cleaning

16

The highlighted data is from stock i, which has data missing for 3 days. *Point 1* is defined as the last adjusted closing price before the missing data. *Point 2* is the fist data point after the missing data. To be remembered is that newer data comes before older data, which means that in the above example, the stock had an adjusted closing price of 15.94 and four days later recorded a 15.84 closing price. The below formula shows how the new data for the three missing points is calculated:

$$data(k,i) = po\text{int}_1 + step*(k-(j-numbers_{mis\sin g})) + sqrt(volatility)*randn$$

where $numbers_{missing}$ represents the number of data points missing, in this case 3, and the formula for *step* is

$$step = \frac{po\text{int}_2 - po\text{int}_1}{numbers_{mis\sin g}}$$

The volatility is calculated from historical prices and then multiplied with a random number generated from a normal distribution with mean 0 and standard deviation 1. This way the randomness of the data is maintained, keeping the overall volatility and, because of the use of the *step*, keeping the overall trend.

Applying this overall process to any missing data or erroneous points, a new data set was generated, a complete data set with no missing points. This set is the one used for all the analysis.

Examples:

Consider the following data, where there is missing history for the first two stocks. This is only part of the data available for each stock. The volatility of Stock 1 before the missing data is of 1.2996 while the volatility of Stock 2 before the missing data is 12.5987.

Stock1	Stock2	Stock3	Stock4
41.77	43.04	100.21	50.34
42.52	44.22	101.33	50.46
41.76	42.09	99.3	48.23
41.52	42.54	100.18	47.89
NaN	43.25	100.04	48.33
41.83	NaN	101.23	48.78
41.66	NaN	102	49.63
42.15	44.23	100.78	49.14
41.65	43.09	100.05	48.04
40.81	42.3	100.71	47.59
40.78	42.2	101.55	48.77
39.7	42.82	100.05	47.35

After implementing the algorithm presented in this section, the following values resulted:

Stock1	Stock2	Stock3	Stock4
41.77	43.04	100.21	50.34
42.52	44.22	101.33	50.46
41.76	42.09	99.3	48.23
41.52	42.54	100.18	47.89
42.11	43.25	100.04	48.33
41.83	**43.6**	101.23	48.78
41.66	**43.13**	102	49.63
42.15	44.23	100.78	49.14
41.65	43.09	100.05	48.04
40.81	42.3	100.71	47.59
40.78	42.2	101.55	48.77
39.7	42.82	100.05	47.35

The new data keeps the volatility and the randomized nature of the rest of the data

3. Trading Strategy

Goal

The purpose of the project was to create a benchmark for trading strategies to be tested on. For exemplification, three trading strategies were chosen: uniformly distributing money over the entire portfolio at the end of every time period, investing in the top n companies($n = 40$ was chosen for exemplification) of the portfolio and using Convex Optimization. A time period is considered either a quarter or a year.

Trading

The system simulates the trading of a portfolio with the help of different portfolio optimization strategies. In order to evaluate the efficiency of the portfolio optimization strategy, the entire data is divided into time periods (quarter or year) and, at the end of each time period, by using the portfolio optimization strategy, a redistribution of the current value of the portfolio is done in accordance with the optimization.

The data used is from the current and previous time periods only, the rest of the data being considered unknown. At the end of each time period all the stocks in which money is invested are sold. The resulting value is reinvested by using the trading strategy of the trader's choice and by only using the "past" data (the data from the current period and the previous periods).

The m day trading period

In order to avoid days where the market had results either above or below average, the decision has been made to invest the money in a period of several days. For

the examples in this paper it was decided to give m a value of 5. This means that the sum that is to be invested in the portfolio is divided equally in 5 and then, on each day, invested in the portfolio according to the trading strategies.

Initial Trading

The first trading period is common for all trading strategies. Starting from an initial investment, say $1,000,000 and the portfolio created in the manner presented above, quarterly and annual investments were made. The same five day trading rule applies here. The initial trading period was chosen to be at the end of 1998, from December 24-th to December 30-st. This way the new quarter and year start from January 1999. The initial investment was divided uniformly in the five days and uniformly within the 170 stocks of the portfolio. This means that each day $200,000 is invested uniformly over the portfolio (the same amount of money is invested in each company, buying as many stocks of each as the money allows).

Further Trading and Trading Strategies

All the data is divided into quarterly and yearly partitions. This data is sent to the Trader.m file to be reinvested. At the end of each time partition a new five day time period starts. During this, the number of stocks invested in each unit is uniformly sold during a period of five days and, in each of these days, the daily resulted money from the sale is reinvested in the portfolio. This creates a 5 day sell-buy period. Using this data and the data at the end of the previous time period (quarter, year), the following tables were generated:

Yearly and quarterly

-         Total Portfolio

-         Total Distribution of Stocks

-         Total Value per Stock

-         Total Gain per Stock

-         Total Percent Gain per Stock

What separates the different trading strategies is the way the money is reinvested in each of the five trading days.

- **Reinvesting Uniformly Over All Stocks**

This method implies uniformly investing each of the five trading day's profit uniformly across the portfolio. At the end of a trading period, on each day of the trading, 1/5 of the stocks is sold and, the total sum resulted from the sale is reinvested uniformly over the 170 stocks at the price of the day. This results in the following portfolio values at the end of each year and quarter:

Yearly Portfolio Value:

Date	Value
12/31/1999	1,284,760
12/29/2000	1,632,686
1/7/2002	2,047,591
1/7/2003	1,964,946
1/7/2004	2,867,571
1/6/2005	3,590,632
1/6/2006	4,375,438
1/9/2007	5,007,642
1/9/2008	5,182,133

Quarterly Portfolio Value:

Date	Value
4/5/1999	1,026,861.35
7/2/1999	1,202,972.41
10/1/1999	1,151,325.59
12/31/1999	1,249,955.89
3/31/2000	1,315,510.88
6/30/2000	1,389,819.57
9/29/2000	1,484,431.99
12/29/2000	1,584,192.95
4/2/2001	1,606,545.48
7/2/2001	1,813,016.45
10/5/2001	1,614,020.22
1/7/2002	1,902,091.99
4/9/2002	2,041,365.53
7/9/2002	1,858,754.26
10/7/2002	1,586,706.55
1/7/2003	1,821,251.24
4/8/2003	1,729,152.93
7/9/2003	2,271,561.69
10/7/2003	2,459,163.69
1/7/2004	2,798,396.71
4/7/2004	2,969,683.03
7/9/2004	2,988,789.06
10/7/2004	3,150,553.36
1/6/2005	3,487,500.34
4/8/2005	3,536,804.58
7/8/2005	3,766,200.33
10/6/2005	3,943,982.18
1/6/2006	4,208,310.96
4/7/2006	4,528,704.25
7/10/2006	4,357,910.71
10/6/2006	4,554,884.41
1/9/2007	4,873,460.53
4/11/2007	5,191,091.5
7/11/2007	5,490,931.8
10/9/2007	5,421,286.23
1/9/2008	4,962,929.55

- **Reinvesting Uniformly Over Top *n* Stocks**

Another strategy is investing uniformly at the end of each time period in the top *n* stocks. The above strategy of uniformly distributing over all stocks is an application of this, with *n = 170.*

This strategy involves calculating the percent return at the end of the current time period (quarter or year) for each stock and singling out the top *n* stocks. The Trader will invest then uniformly in these *n* stocks following the same 5 day trading strategy. For exemplification of this strategy, *n* was chosen to be 40. The following portfolio values resulted at the end of every quarter and year:

Yearly Portfolio Value:

Date	Value
12/31/1999	1,284,760.23
12/29/2000	1,537,435.5
1/7/2002	1,708,320.76
1/7/2003	1,932,339.27
1/7/2004	2,715,143.56
1/6/2005	3,651,453.93
1/6/2006	4,982,715.92
1/9/2007	5,370,633.99
1/9/2008	5,776,921.03

Quarterly Portfolio Value:

Date	Value
4/5/1999	1,026,861.35
7/2/1999	1,180,977.1
10/1/1999	1,164,115.36
12/31/1999	1,502,352.24
3/31/2000	1,682,604.51
6/30/2000	1,764,676.51
9/29/2000	2,024,598
12/29/2000	2,040,107.08
4/2/2001	2,007,561.01
7/2/2001	2,338,822.37

10/5/2001	1,884,074.7
1/7/2002	1,982,570.38
4/9/2002	2,242,480.34
7/9/2002	2,210,408.77
10/7/2002	1,949,542.77
1/7/2003	2,059,621.49
4/8/2003	1,967,801.63
7/9/2003	2,397,642.88
10/7/2003	2,763,322.83
1/7/2004	3,185,694.03
4/7/2004	3,345,854.13
7/9/2004	3,378,322.98
10/7/2004	3,551,027.48
1/6/2005	4,069,085.28
4/8/2005	4,199,760.16
7/8/2005	4,384,896.42
10/6/2005	4,852,948.23
1/6/2006	5,131,824.16
4/7/2006	5,869,688.46
7/10/2006	5,591,078.66
10/6/2006	5,438,122.43
1/9/2007	5,709,734.48
4/11/2007	6,142,800.26
7/11/2007	6,790,026.09
10/9/2007	7,020,910.88
1/9/2008	6,712,656.47

- **Markowitz Portfolio Optimization**

A third trading strategy tested on the system was the Markowitz Portfolio Optimization. This method minimizes the return variance (associated with the risk of the portfolio) subject to achieving a minimum acceptable mean return. No short positions are accepted in this implementation, but this feature can be easily added.

At the end of each time period the program calculates the *matrix of daily returns* of each stock in the time period. This is calculated using the formula:

$$daily_{return} = \frac{price_{today} - price_{yesterday}}{price_{yesterday}}$$

The last row is discarded.

The vector $\bar{p}$ is calculated, where $\bar{p}$ is the vector of the mean of the daily returns for each stock. The next step is calculating the *matrix of daily centered returns* (P) by subtracting from the *matrix of daily returns* the respective average from the vector. This means that for stock i, we will subtract from the entire i'th column of the *matrix of daily returns* the i'th element if $\bar{p}$ ($\bar{p}$ [i]).

The covariance of P, the *matrix of centered daily returns* can be calculated by either using the built in Matlab function *cov(Matrix)* or by using the formula:

$$\Sigma = \frac{1}{columnNumber(P)} * P^T \cdot P$$

The Trader uses quadratic optimization (the Matlab function quadprog()) to

$$\text{Minimize } x^T \Sigma x$$

$$\text{Subject to } \bar{p}^T x \geq r_{min}$$

$$1^T x = 1$$

$$lb \leq x \leq ub$$

starting from x0, where x0 is the previous distribution.

The r_{min} above was chosen as 0.0001 and 0.0015. The lower boundary (lb) was chosen to be 0 and the upper boundary (ub) was taken to equal 0.025. The lower boundary was so chosen to prevent overinvestment in one asset. This way at least 40 stocks will be invested in.

The following portfolio values were obtained for the yearly and quarterly time period:

25

Yearly Portfolio Value:

Date	Value $r_{min}=0.0001$	Value $r_{min}=0.0015$
12/31/1999	1,284,760	1,284,760
12/29/2000	1,483,025	1,468,114
1/7/2002	1,986,954	1,939,264
1/7/2003	2,281,612	2,270,505
1/7/2004	2,755,897	2,983,702
1/6/2005	3,673,209	4,038,005
1/6/2006	4,366,913	4,990,428
1/9/2007	4,980,195	5,489,648
1/9/2008	5,043,998	5,646,559

Quarterly Portfolio Value:

Date	Value $r_{min}=0.0001$	Value $r_{min}=0.0015$
4/5/1999	1,026,861	1,026,861
7/2/1999	1,170,161	1,182,884
10/1/1999	1,131,227	1,144,729
12/31/1999	1,155,527	1,276,501
3/31/2000	1,208,470	1,317,036
6/30/2000	1,246,045	1,361,044
9/29/2000	1,313,691	1,422,917
12/29/2000	1,397,930	1,512,026
4/2/2001	1,444,735	1,559,966
7/2/2001	1,591,195	1,725,698
10/5/2001	1,577,573	1,712,257
1/7/2002	1,706,924	1,828,181
4/9/2002	1,865,212	1,998,096
7/9/2002	1,835,137	1,972,306
10/7/2002	1,735,465	1,753,527
1/7/2003	1,865,153	1,836,435
4/8/2003	1,801,853	1,801,671
7/9/2003	2,124,413	2,052,560
10/7/2003	2,286,382	2,170,045
1/7/2004	2,539,765	2,408,587
4/7/2004	2,730,409	2,584,606
7/9/2004	2,804,207	2,633,172
10/7/2004	2,753,587	2,634,782
1/6/2005	2,976,855	2,855,994
4/8/2005	3,130,753	2,887,313

7/8/2005	3,238,100	2,948,680
10/6/2005	3,242,820	3,023,356
1/6/2006	3,376,832	3,132,960
4/7/2006	3,255,932	3,258,671
7/10/2006	3,146,151	3,102,340
10/6/2006	3,404,975	3,223,487
1/9/2007	3,597,615	3,392,846
4/11/2007	3,785,716	3,569,433
7/11/2007	3,893,868	3,684,174
10/9/2007	3,845,893	3,732,415
1/9/2008	3,619,739	3,535,405

4. Conclusions

In order to compare the different trading strategies, one must analyze what the value of the portfolio in time and its standard deviation (representing the risk of the portfolio) are for the different trading approaches. The comparison was done both yearly and quarterly. The Markowitz Optimization was taken with the minimum return of 0.0015.

Yearly Trading
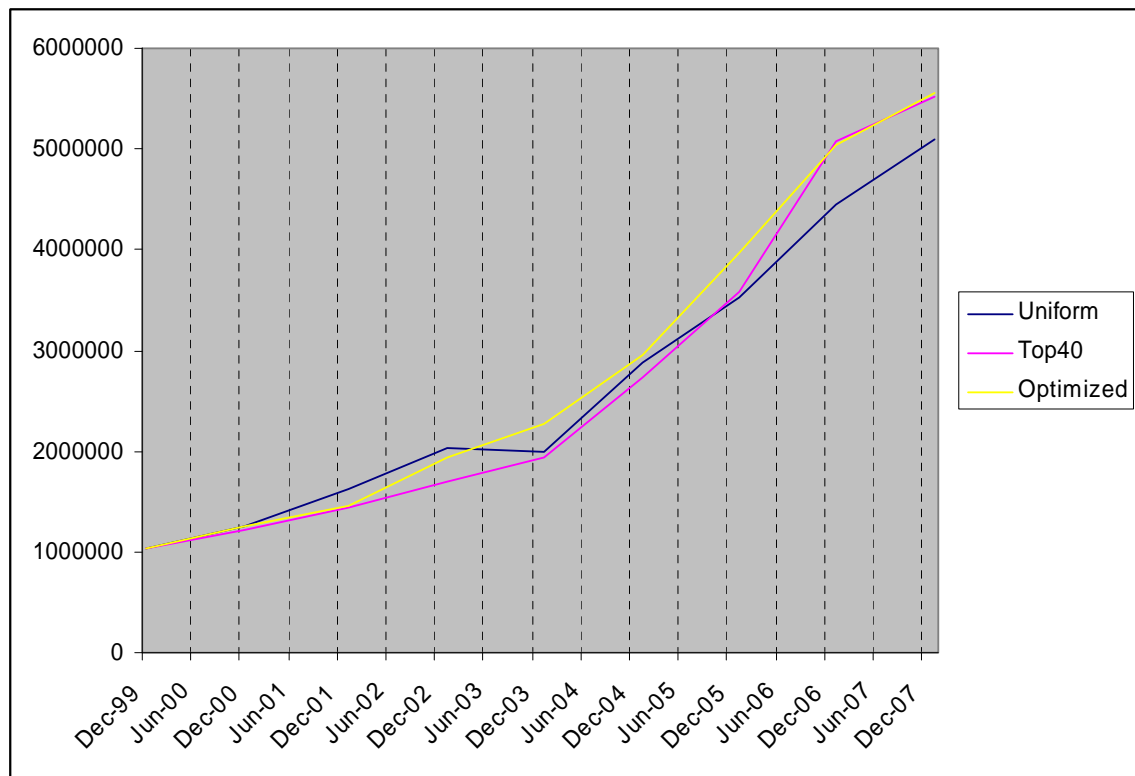
Figure 3. Yearly Portfolio Values

Since the data used spans only over nine years and trading is performed at the end of each year, there are only nine implementations for each strategy. Through the year the distribution of stocks in the portfolio remains unchanged. Figure 3 represents the yearly portfolio values of each stock. For this plot the value of the portfolio was calculated

28

monthly. Both the Optimized and the Top40 strategies have similar returns at the end of the nine year period. On the other hand, the Optimized distribution of stocks generated a higher average portfolio value. The uniform distribution was the least successful strategy, generating overall the smallest portfolio value.

Figure 4 represents the standard deviation of each strategy when it was implemented yearly. It is a representation of the risk of each trading strategy. Since the Markowitz Optimization strategy is designed to minimize risk while achieving a minimum return, it was most likely to have the lowest overall standard deviation. As seen in the below figure, the optimized strategy had the lowest standard deviation, while the Top 40 strategy was overall the riskiest. The reason why the Top40 strategy is the riskiest is because of limited diversification. The other strategies distribute the money over more stocks, creating hedging.
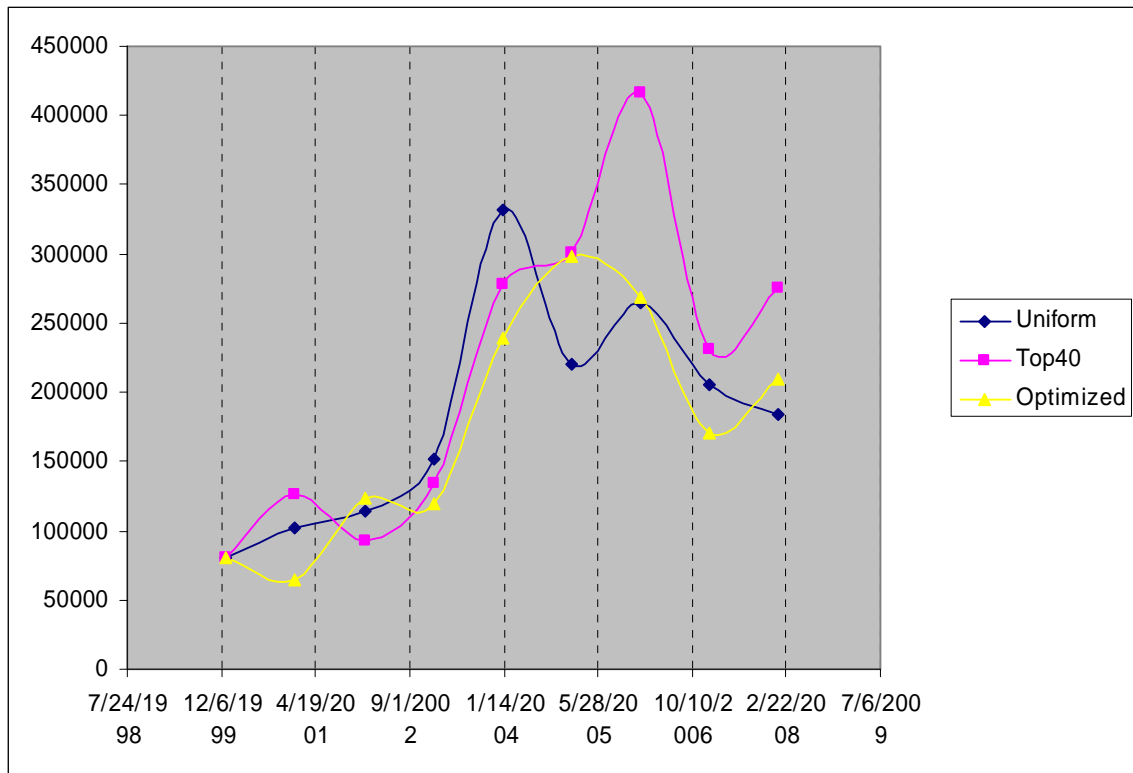
Figure 4. Yearly Standard Deviation

Quarterly Trading

The quarterly distribution involves 36 different instances of implementation of each strategy. On the other hand, since most of the trading decisions are made by taking into consideration the data from the current quarter and the previous one, there is less data to base the decision on.

Figure 5 represents the portfolio values when trading has been done at the end of each quarter. The top 40 strategy was throughout the 9 years the one with the highest returns. The Markowitz Optimization had the lowest returns.
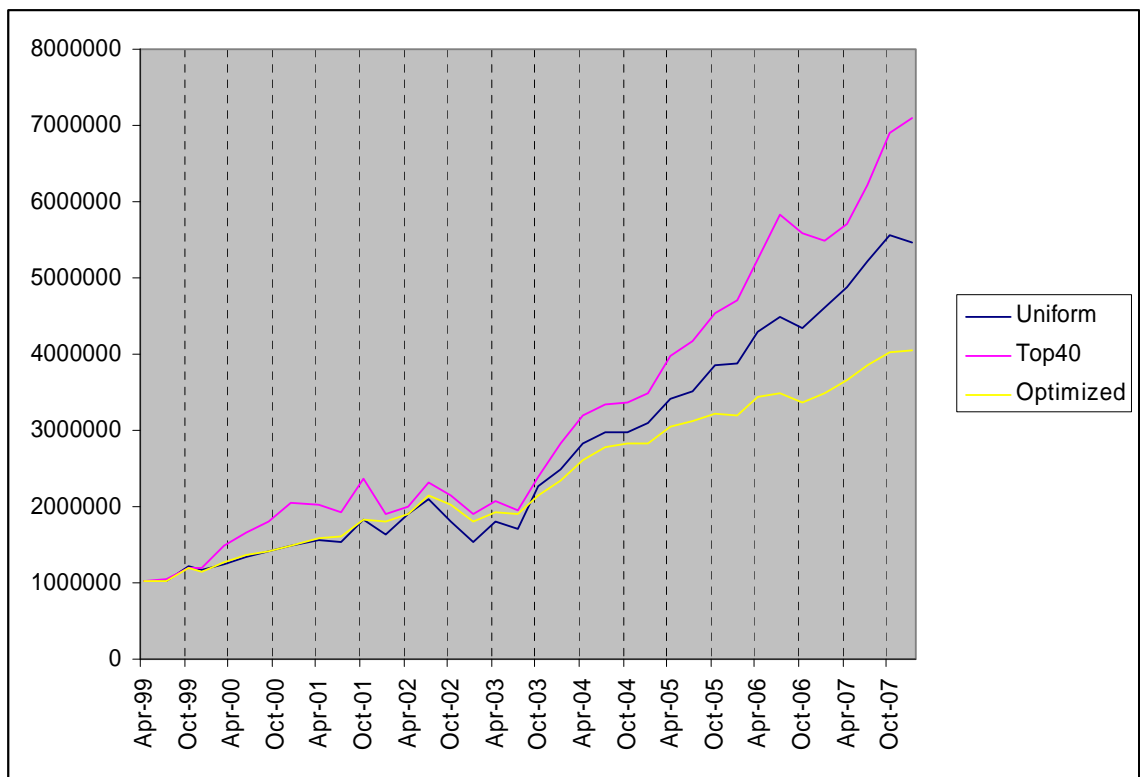

Figure 5.Quaterly Portfolio Values

As seen in Figure 6, which represents the standard deviation of each strategy, the Markowitz Optimization has the lowest risk of all. The riskiest strategy remains the Top40 one.
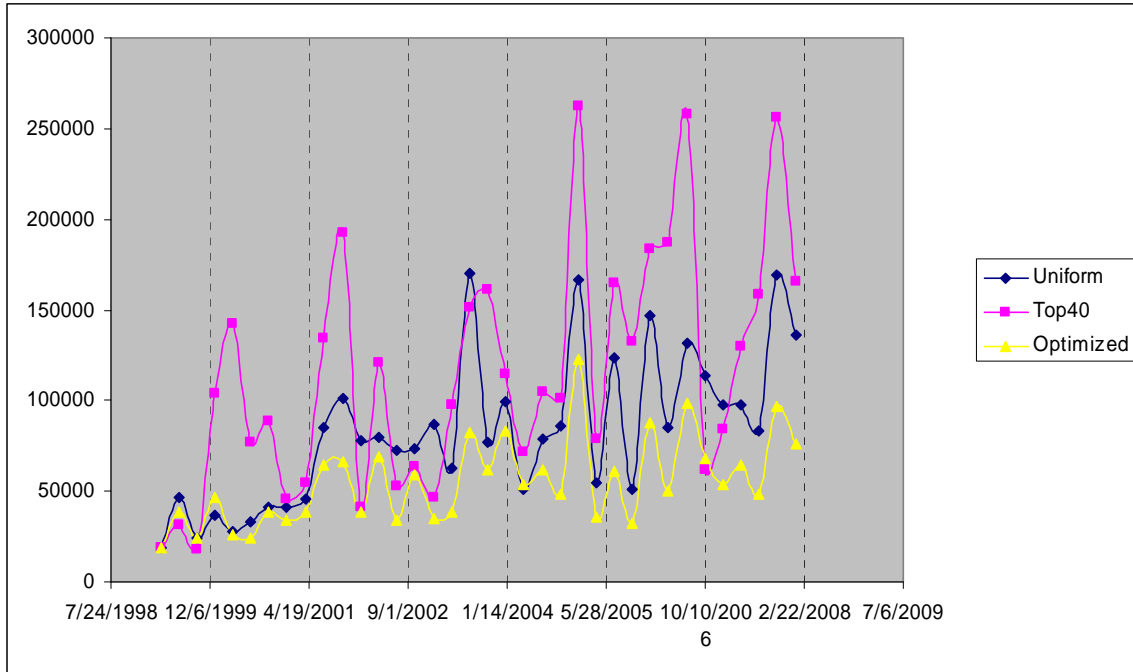
Figure 6. Quarterly Standard deviation

Final Conclusions

This system can be used to produce the data needed to compare trading strategies. The results reached by analyzing the data generated by the three implemented strategies are consistent with the fact that portfolios with lower risk generate lower return, while portfolios with higher risk generate higher return. Investing quarterly resulted in generating on average the highest returns since it allowed strategies to respond to the market and better correct themselves.

This project could be extended by adding other trading strategies. Better comparison tools for analyzing the differences between trading strategies could be added, Also, the algorithm could be tested on a larger set of stocks.

31

Annex 1 – Tables and Matlab code – The Data tables and the Matlab code is available and has been submitted separately.