Portfolio Modeling, Analysis and Management

by

Divakar Chitturi

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved October 2010 by the
Graduate Supervisory Committee:

Armando Rodriguez, Chair
Konstantinos S. Tsakalis
Jennie Si

ARIZONA STATE UNIVERSITY

December 2010

ABSTRACT

A systematic top down approach to minimize risk and maximize the profits of an investment over a given period of time is proposed. Macroeconomic factors such as Gross Domestic Product (GDP), Consumer Price Index (CPI), Outstanding Consumer Credit, Industrial Production Index, Money Supply (MS), Unemployment Rate, and Ten-Year Treasury are used to predict/estimate asset (sector ETF's) returns. Fundamental ratios of individual stocks are used to predict the stock returns. An a priori known cash-flow sequence is assumed available for investment. Given the importance of sector performance on stock performance, sector based Exchange Traded Funds (ETFs) for the S&P and Dow Jones are considered and wealth is allocated. Mean variance optimization with risk and return constraints are used to distribute the wealth in individual sectors among the selected stocks. The results presented should be viewed as providing an outer control/decision loop generating sector target allocations that will ultimately drive an inner control/decision loop focusing on stock selection. Receding horizon control (RHC) ideas are exploited to pose and solve two relevant constrained optimization problems. First, the classic problem of wealth maximization subject to risk constraints (as measured by a metric on the covariance matrices) is considered. Special consideration is given to an optimization problem that attempts to minimize the peak risk over the prediction horizon, while trying to track a wealth objective. It is concluded that this approach may be particularly beneficial during downturns - appreciably limiting downside during downturns while providing most of the upside during upturns. Investment in stocks during upturns and in sector ETF's during downturns is profitable.

To the memories

ACKNOWLEDGMENTS

I am very grateful for the cooperation and support of my advisor Dr. A. A. Rodriguez, who has shown a great deal of patience and confidence in my work.

Besides my advisor, I would like to thank Srikanth Sridharan for his help and guidance.

I would like to acknowledge the help of Lakshmi Gade and Vishnu Sudha. There are several others who have contributed to my success. In particular N. Srinivas, Brent Mashburn, Chandra Geethika, Dr. Deepika Sidhu, Arun Rambhatla, Donald Hutchins and Venkata Gudipudi.

Finally I would like to thank my family for their consistent support and encouragement.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

1 INTRODUCTION

The art of making investment decisions to make the largest possible return is Portfolio Management. Active portfolio management and passive portfolio management have been widely used since early 19th century. The studies presented in this thesis, are intended to investigate different models and methods to achieve an investor objectives to maximize his/her investments while minimizing the risk.

1.1 *Motivation*

The motivation behind the thesis is to work towards development, of an environment which enables systematic approach for data acquisition, portfolio components (macro, sector, stocks) analysis, modeling, prediction and management (decision making). Focus of this thesis is limited to stocks and sector ETF's with the goal of incorporating Macro indicators, business cycle analysis, sector ETF's, fundamental ratios and technical indicators.

Various studies (Chen 1986), (Boyd 2001), (O. Lamont 2001) have been performed to determine the impact of macroeconomic and global factors on stocks returns. Similarly importance of sector/ industry allocation compared to the country allocation was studied by (King 1966), (Livingston 1977). Fundamental and technical analysis techniques have been used to predict the stock returns (Fama 1981) (French 1992). Recently, there has been an increasing interest in applying control methodologies to the problem of portfolio optimization. In (B. Durtschi 2009), the authors discuss the use of portfolio Optimization problems as a platform for introducing systems theory to students. Options hedging (J. Primbs 2010), index tracking (Sung 2008), (Zhou 2006), asset liability management (F.Herzog 2007), and constrained

wealth maximization (S. K. F. Herzog 2006) are a few of the common financial problems to which control concepts have been applied. Some of the control methodologies that have been used include linear matrix inequalities (Paiva 2002), linear quadratic regulation (Yin 2003) (G. D. F.Herzog 2007), and model predictive control (Sung 2008), (F. Herzog 2006). This thesis addresses the problem of asset allocation by a risk-conscious investor. A systematic top down approach is used which considers macroeconomic, global and industry factors along with the individual stock fundamentals to determine the asset allocation. A receding horizon (RH) approach is used in order to meet the objectives and constraints.

1.2 *Contributions*

This thesis forms a good starting point for development of environment which facilitates macro, sector, and stock data acquisition, analysis, modeling, prediction and portfolio management. It helps in hierarchical decision making in the presence of uncertainty. Also permits easy comparison of different scenarios. A unique top down approach for investment has been explored with a fundamental belief of macroeconomic factors influencing the sector performance which in turn influences the individual stock performance. The framework includes macro, and asset modeling. Different models including Factor models, regression models, hybrid models and state space models for asset return forecasts are considered. Different methodologies including regression, recursive least square, moving window and fixed window methods are explored for accurate forecasting of the macro variable data, and asset returns. Key fundamental ratios of more than 1000 companies are analyzed which can be used in the asset models. An alternative to the classical wealth

maximization approach to portfolio optimization is presented. The effect of different risk (as measured by covariance matrices) tolerance on the portfolio performance is considered. Apart from risk constrained wealth maximization, a mini-max risk subject to wealth tracking constraints approach might be used by investors to attain their objectives is showed. A pictorial representation of the integrate hierarchical portfolio management approach is shown in Fig 1.

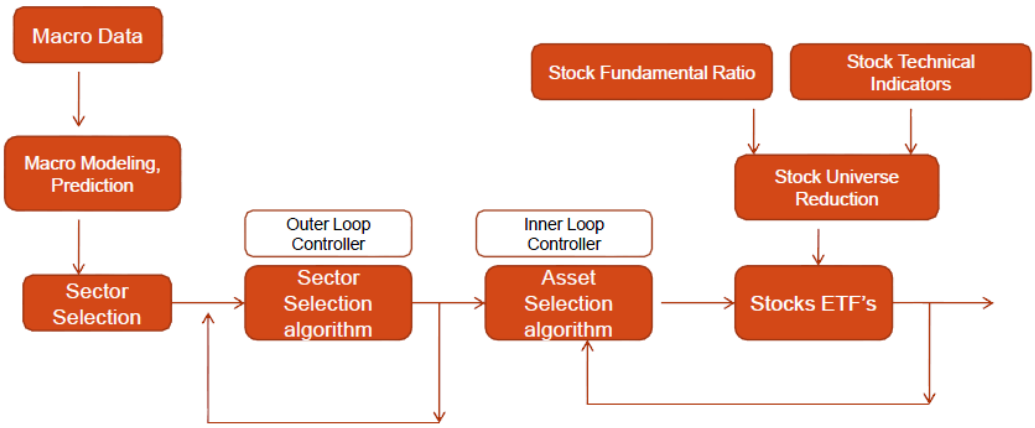

Fig 1: Integrated Hierarchical Portfolio management

1.3 *Challenges*

Indentifying the impact of low frequency macroeconomic data on high frequency stock data and incorporate this knowledge to predict the asset returns is a significant challenge. Having known that sector performance plays an important role in portfolios, categorizing, identifying and selection of crucial sectors for asset returns forecasts requires understanding of the sectors and depends on identifying the trends of the economy to a great extent. Time series forecasting models have often been used to forecast stock prices (Makridakis 1982), (McKenzie 1985), (Shephard 1993) and (Winker

2003). Recently, more sophisticated techniques such as support vector machines (V. Hanser 2006), (Batchelor 2000), machine learning algorithms (Deng 2002), (D. K. Roley 1985) etc. have been used in trying to identify stocks that are likely to perform significantly better than the market. The effect of news on stock prices has been considered in (Best 1991). Apart from portfolio optimization, asset valuation plays a significant role in derivatives pricing as well. Errors in asset valuation can thus have significant impact on hedging strategies. For certain types of optimizations, small variations in asset return predications are likely to result in significant changes to the allocations. Simple mean-variance optimizations can suffer from this effect (Doege 2005). Several modifications, such as the Black-Litterman model, attempt to reduce this sensitivity to asset prices. Estimating and predicting risk is also an active area of research. Several different measures for risk exist, such as return covariance matrices, value-at-risk (VaR), conditional VaR (CVaR), downside risk etc. Convex risk measures are attractive due to their favorable computational properties (P. Samuelson 1967). Coherent risk measures encourage diversification ((Markowitz 1952) discusses the benefits of diversification); however, some popular risk-measures are not coherent (e.g. VaR).

1.4 *Literature review*

1.4.1 Macroeconomic analysis

Many studies have been performed to confirm that there is an influence of macroeconomic factors on the stock returns. Some of the popular works are,

- (Chen 1986) examined equity returns relative to a set of macroeconomic variables and suggested that Growth in industrial

production, Changes in the risk premium, Twists in the yield curve, Measures of unanticipated inflation, Changes in expected inflation during periods of volatile inflation were important in explaining past stock market returns.

- (Boyd 2001) Found that macro news has time varying effects on the stock returns. They also conclude that, during an economic expansion high unemployment, raises the stock prices, but lowers the stock value during economic contraction.
- (Lamont 2001) Tries to identify priced macro factors, by determining whether a portfolio constructed to track the future path of macro series earns positive abnormal returns. He concludes that PF that tracks the industrial production, consumption and labor income gives abnormal positive returns, while the portfolio tracking the CPI does not.
- (Fama 1981) and (Schewrt 1981) attempts to explain anomalous stock return-inflation relations.
- (Castanias 1979) Reports that the variance of stock prices rises around the days of most economic news events which he interprets as a reflection of new information appearing.
- (Schewrt 1981) Examines the stock market reaction to the monthly CPI inflation rate announcement and does use a measure of unexpected inflation rather than just the announced rate.
- Aggregate stock returns are negatively related to inflation and to money growth - (Fama 1981), (Roll 1983), (Roley 1985) , (Bodie 1976).

- Macro economic factors affect the performance of the stocks to a greater extent and findings thus far provide evidence that macroeconomic fundamentals are indeed an important driving force behind financial market returns (Victor Fang 2009).
- There have been elaborate studies on the impact of the economic factors. Some of them are a) (Fleming 1997), (Bollerslev 2000) and (Green 2004) find that news surprises from GDP, inflation rate, unemployment rate, and consumer confidence are related to changes in Treasury yields especially around the time of the announcements.
- (A.Protopapadakis 2002) Establish the effect of macroeconomic news on the volatility of asset returns.
- Stock return variability was unusually high during the 1929-1939 great depression (Officer 1973)

1.4.2 Sector Analysis

Studies have been performed to determine the importance of sector analysis.

- The movement of a group of security price changes can be broken down into market and industry components (King 1966)
- Industry allocation is an increasingly important consideration for active managers of global equity portfolios .The return on security I is affected by the global factor, the industry, the country to which the stock belongs, and by an idiosyncratic disturbance. The industry factors have become an increasingly important component of security returns. More importantly, diversification across industries now provides greater risk reduction than diversification across countries (Stefano Cavaglia 2000)

- (Anne Sophie E Vanroyen 2002) Measure the relative importance of country and sector effects in these factors, and find that sectors have become as important as countries since October 2000.
- Scenario models provide an alternative to factor models, wherein various future scenarios (and their probability of occurrence) are considered, and the performance of each security under each scenario is evaluated (B. I. Jacobs 2005). This is used to arrive at an estimate of future performance of the portfolio as a whole.
- The authors of (A. R. Chen 1986) examined equity returns relative to a set of macroeconomic variables and suggested that growth in industrial production, changes in the risk premium, twists in the yield curve, measures of unanticipated inflation and changes in expected inflation during periods of volatile inflation were important in explaining past stock market returns.

1.4.3 Stock Analysis

Fundamental ratios are widely used in the stock price prediction. (French 1992) determines the significant relation between returns and variables like size, price-to-book ratio and conclude that the earnings-price ratio is significant when it is the unique explaining variable for the cross-section of stock returns. However, its significance disappears when book-to-market ratio is also taken into account. (Jefrey Pontiff 1998), (Shanken 1997) discuss about the predictive ability of book-to-market ratio. (EF Fama 1988) & (Shiller 1988) have documented that, dividend yield predicts stock returns with some success. While (O. Lamont 1998) argues that, the P/E ratio has independent predictive power for excess returns in addition to the dividend-price ratio.

(Lewelle 1999) Book-to-market ratio predicts economically and statistically significant time-variation in expected stock returns. (Jeffrey Jaffe 1989) analyses the relation between stock returns and the effects of size and earnings-to-price ratio.

1.4.4 Portfolio Optimization

Markowitz considered the problem of single-period portfolio optimization. The transition to multi-period optimization and dynamic models can be found in (Samuelson 1969.) and (Merton 1969). Some early optimization techniques include stochastic dynamic programming and solving the Hamilton-Jacobi-Bellman (HJB) equations. These solutions are difficult to compute, especially when the dimension of the problem increases (Bertsekas 2005) (G. D. F.Herzog 2007). Necessary and sufficient conditions for the existence of an optimal portfolio have been considered in (D. P. Bertsekas 1974) . (H. Peyrl 2005) and (Z. Chen 2008) considers some of the numerical challenges associated with stochastic control problems in financial applications. In (G. D. F.Herzog 2007), the authors prove that a model predictive control (MPC) solution is suboptimal (but at least as good as a pure-open loop solution), and they discuss obtaining bounds on the relative error between these techniques. RHC methods, however, have the advantage of being computationally feasible and capable of handling complex constraints. A survey on MPC techniques can be found in (Jalali 2006). (M. V. Kothare 1996) and (Smith 2004) discuss the problem of robust MPC using linear matrix inequalities. Robust minimax optimization is discussed in (Witsenhausen 1966), while (Lofberg 2003) uses it in an MPC framework. Solving constrained and uncertain MPC problems has also been discussed in (Richards 2002).

2 PORTFOLIO MANAGEMENT: MODELLING REQUIREMENTS

In this chapter each and every component used in the thesis are introduced and are explained with their definitions and their usage in the integrated framework.

2.1 *Asset Classes and Allocation Issues*

Asset class analysis, involves identifying the right asset for investment. Assets classes worth investigating are stocks, bonds, real estate, currency, commodities and etc. In this thesis, equities are focused. Other asset classes are equally attractive and can be used to diversify investments or can be used for hedging. Introduction to types of assets is given below

2.1.1 Equities

Equities - are defined as the ownership in any asset after all the debts associated with the asset are paid off. This is called equity, which can be sold to make it as cash instantly (Investopedia, Investopedia n.d.). Stock can be called as equity since it represents ownership in a company, this stock can be sold readily for cash. Equities are primarily focused, since they are the most popular and widely invested asset class and something which can be easily understood by any common investor when compared with the other asset classes. One of the other reasons equities are focused is, all/ most of the information with respect to the individual companies, stock can be easily obtained from the respective company websites, or from the internet

2.1.2 Derivatives

Securities which provide payoffs depending on or are contingent on the values of other assets such as commodity prices, bond and stock prices, or market index values. Some of the examples are futures and options.

Futures contracts oblige traders to purchase or sell an asset at an agreed price on a specified future date. Options are classified into call and put options where the call option gives its holder the right to purchase an asset at an agreed price on or before a specific expiration date. Put option gives its holder the right to sell an asset for a specific price on or before the expiration date (Frank K. Reilly 2002)

2.1.3 Mutual Funds

Mutual funds in simple terms are collective investment scheme, where a person/organization, collects money from the investors, and invests them in different assets such as stocks, government bonds, currencies, commodities and etc. The gains obtained from the investment are distributed to all the investors who contributed to the fund. These funds are professionally managed by investment gurus based on their knowledge on markets.

One can buy into the funds at a specific date, or anytime during the trading hours depending on the type of fund one is investing. Investing in a fund means buying/selling a unit of fund whose value is current market value of funds holding minus the funds liabilities. This value is called as net asset value represented in per share basis.

Security Exchange commission hereafter referred as SEC defines a mutual fund as a company that pools money from many investors and invests the money in stocks, bonds, short-term money-market instruments, or other securities. (SEC n.d.) Focus on mutual funds is limited as equities are explored in details.

2.1.4 ETF

As defined by SEC (SEC n.d.), Exchange traded funds can be understood as a security which tracks the index, such as SP500, Dow Industrial etc. This can be bought and sold like a regular stock whose value is determined by the demand /supply of the security. (Investopedia n.d.)

One of the greatest advantages of investing in ETF is that, it gives diversification of the investment, since the index does not represent one company in specific. Index represents a basket of companies across all sectors, thereby reducing our exposure to the risks associated with one particular company and or sector.

In this thesis, ETF's are used to a greater extent for the analysis of sector performance. Sector ETF's represent the whole sector rather than one or two leading companies. Widely used ETF's are SPDRs' which are managed by State Street Global advisors. Every sector has an ETF representing all the companies in that particular sector. Apart from this there are also index tracking ETFs like SPY which track the SP500.

2.1.5 Commodities

A basic good used in commerce that is interchangeable with other commodities of the same type. They are most often used as inputs in the production of other goods or services. The sale and purchase of commodities is usually carried out through futures contracts on exchanges that standardize the quantity and minimum quality of the commodity being traded. (Investopedia n.d.). Commodities can be used as hedging instruments, to make a profit during high inflation.

2.1.6 Real Estate

As defined by SEC (SEC n.d.), Real Estate investment trusts are the instruments which invest in real estate or real estate related assets, including shopping centers, office building, hotels and etc. Focus on real estate funds is extremely limited.

2.1.7 Emerging Markets

Markets of the developing nations such as BRIC (Brazil, Russia, India, and China) where the rate of growth is significantly higher than developed countries such as US, Europe are referred as Emerging markets. If one is interested in investing in other countries, then he or she needs to purchase the ETF's which track the country's Main Index eg, BSE SENSEX of India. Sometimes shares of the companies based out and operating in other countries can be bought and sold in the local stock market, if they are listed as ADR's. Emerging markets have great growth rate and hence investing in them is a wise option.

2.1.8 Treasuries

Treasury securities—including Treasury bills, notes, and bonds—are debt obligations issued by the U.S. Department of the Treasury. Treasury securities are considered one of the safest investments because they are backed by the full faith and credit of the U.S. government. (SEC n.d.). Treasuries are used for reducing the risk of our investment since it is assured return from the government. Treasuries can be used as a tool to minimize risk in the portfolio optimization.

2.2 *Macroeconomic View: Sector Based Business Cycle Analysis*

Below are the key macro economic variables which are widely used in policy decisions, economic growth calculations and most important, they are used to understand the health of the economy and economic state of the people.

2.2.1 GDP

It represents the total dollar value of all goods and services produced over a specific time period. It is one the primary indicators used to gauge the health of a country's economy. (Investopedia n.d.)

"GDP is the value of all goods and services produced in the country without regard to its allocation among domestic and foreign claims **"**(Janice Peterson 1999)**.**

"Gross Domestic Product (GDP) is a measure of the total domestic economic activity. It is the sum of all incomes earned by the production of goods and services on UK economic territory, wherever the earner of the income may reside. GDP is equivalent to the value added to the economy by this activity. Value added can be defined as income less intermediate costs. Therefore growth in GDP reflects both growth in the economy and price changes (inflation)". (Treasury n.d.)

GDP = Consumption + Investment + Government spending + Exports – Imports. (Carlos M. Gutierrez 2007)

2.2.2 Real GDP

Real gross domestic product -- the output of goods and services produced by labor and property located in the United States (Bureau of Economic Analysis n.d.) . This inflation-adjusted measure that reflects the value of all goods and services produced in a given year, expressed in base-year prices. Often

referred to as "constant-price", "inflation-corrected" GDP or "constant dollar GDP" (Investopedia n.d.). Gross domestic product per capita adjusted for differences in the purchasing power of currencies (Irving B. Kravis 1978). In this thesis, GDP is used for the macro economic analysis, sector analysis, and impact of GDP on stocks. GDP is also forecasted into future, to determine/ predict the performance of the stocks which would help in identifying the stocks which would perform better.

2.2.3 Unemployment

The unemployment rate represents the number unemployed as a percent of the labor force (The labor force is the sum of employed and unemployed persons).

Persons are classified as unemployed if they do not have a job, have actively looked for work in the prior 4 weeks, and are currently available for work. Persons who were not working and were waiting to be recalled to a job from which they had been temporarily laid off are also included as unemployed. Receiving benefits from the Unemployment Insurance (UI) program has no bearing on whether a person is classified as unemployed (Bureau of labour statistics n.d.) .In this thesis, we use unemployment data for macro economic analysis and also sector analysis. We try to find out the impact of unemployment over the stock price (SP500 index in particular)

2.2.4 Inflation

A substantial rise of prices caused by an undue expansion in paper money or bank credit (Hazlitt n.d.)

2.2.4.1 CPI

The Consumer Price Index (CPI) is the prices paid by urban consumers for a representative basket of goods and services (Bureau of labour statistics n.d.). Consumer price indexes often are used to escalate or adjust payments for rents, wages, alimony, child support and other obligations that may be affected by changes in the cost of living.

A measure that examines the weighted average of prices of a basket of consumer goods and services, such as transportation, food and medical care. The CPI is calculated by taking price changes for each item in the predetermined basket of goods and averaging them; the goods are weighted according to their importance. Changes in CPI are used to assess price changes associated with the cost of living (Investopedia n.d.). CPI is one of the most frequently used statistics for identifying periods of inflation or deflation. This is because large rises in CPI during a short period of time typically denote periods of inflation and large drops in CPI during a short period of time usually mark periods of deflation

2.2.4.2 PPI

The Producer Price Indexes (PPIs) are a family of indexes that measure changes in the selling prices received by domestic producers of goods and services. They formerly were referred to as Wholesale Price Index (Bureau of labour statistics n.d.).

A family of indexes that measures the average change in selling prices received by domestic producers of goods and services over time. PPIs measure price change from the perspective of the seller (Investopedia n.d.)

2.2.5 Money Supply

The entire quantity of bills, coins, loans, credit and other liquid instruments in a country's economy (Investopedia n.d.). It is further classified into M0, M1, M2 and M3. **M0** is a measure of the money supply which combines any liquid or cash assets held within a central bank and the amount of physical currency circulating in the economy. M0 (M-zero) is the most liquid measure of the money supply. It only includes cash or assets that could quickly be converted into currency. **M1** is a category of the money supply that includes all physical money such as coins and currency; it also includes demand deposits, which are checking accounts, and Negotiable Order of Withdrawal (NOW) Accounts. **M2** is a category within the money supply that includes M1 in addition to all time-related deposits, savings deposits, and non-institutional money-market funds. **M3** is the category of the money supply that includes M2 as well as all large time deposits, institutional money-market funds, short-term repurchase agreements, along with other larger liquid assets. In this thesis, we use M2 money supply data, since it indicates all the savings and deposits, including the M1 data. Since M3 has short term funds included, that might not be a good indicator for mid to long term investment.

2.2.6 Interest rates

This is the rate at which the borrower pays the interest to the lender for the amount he borrows. Interest rates targets are also a vital tool of monetary policy and are taken into account when dealing with variables like investment, inflation, and unemployment. (Wikipedia n.d.)

2.2.7 Housing Data

Housings data is a monthly statistic released by the government, indicating the number of housing units available, growth in the housing sector, growth in demand for new houses and etc. These numbers drive all the auxiliary industries which are based on housing and infrastructure development.

Housing and real estate have gathered lots of attention both from media and the public after the 2008 economic crash, which was primarily caused by the home loan mortgages. To sum up the whole problem, the home loans were securitized and sold to multiple countries and investors. Rise in demand in these, led to distributing the loans to non eligible lenders who ultimately ended up defaulting on the loans, which directly caused excessive losses for the loan distributing companies as the rate of defaulters increased day by day. Real estate is an attractive investment option however we believe focus on this requires more time and data to come up with an sound investing methodology.

2.2.8 Trade Deficit

An economic measure of a negative balance of trade in which a country's imports exceeds its exports. A trade deficit represents an outflow of domestic currency to foreign markets. The value of all the goods and services that a country sells to other countries (exports) minus the value of all the goods and services the same country buys from foreigners (imports) is called our **trade balance.** If the value of the trade balance is positive, we have a **trade surplus** and we export more than we import (in dollar terms). A **trade deficit** is just the opposite; it occurs when the trade balance is negative and the value of what we import is more than the value of what we export (Robin Bade 2003). Trade deficit may not be a leading or lagging indicator of the day

to day economic health of the country. However it is a very important parameter which has to be taken care.

2.2.9 Savings Data

Savings - as the name suggests is the amount of money the government has, as a result of savings from its people. Savings are the key driver of liquidity. Banks get cash for all its operations only when its customers deposits money for a agreeable rate of interest. If the banks don't have cash, the whole cycle of investment is disrupted. At these times, government may decide to increase the liquidity in the system, by increasing the savings interest rate so that people decide to save their earnings for a higher interest, stopping the individuals and industries in withdrawing the invested cash for a specific period of time, or decide to print more currency. Interesting questions such as why can't government print more money come up. However these action result in other problems such as inflation.

2.2.10 Consumer Spending

It is defined as the amount of money consumer spends for his or her requirements; it is also terms as personal consumption expenditure. Consumer spending is the key indicator of the health of the economy.

2.2.11 Other Leading and Lagging Indicators

2.2.11.1Recession

A recession is the period when overall economic activity is actually declining—and production, employment, and sales are falling—rather than just anemic or below normal (Lansing 2003). A significant decline in activity spread across

the economy, lasting more than a few months, visible in industrial production, employment, real income, and wholesale and retail trade. Popular rule of thumb is that two consecutive quarterly declines in real GDP signal a recession.

Some of the other key economic of recession are, Real GDP, Real House hold spending (defined as real personal consumption expenditures plus real residential investment) since this category of spending accounts for about three-fourths of U.S.GDP. It is observed during previous recessions when household spending typically slowed prior to the business cycle peak and then declined for two or three quarters. Strong performance of household spending during the past two years. Fiscal stimulus in the form of tax rebates, cuts in marginal tax rates, and extended unemployment benefits provided support to consumer disposable income. Attractive financing deals offered by domestic auto manufacturers gave a significant boost to consumer durables purchases. Most importantly, low mortgage interest rates spurred record home sales and set off a refinancing boom that allowed consumers to tap the equity in their homes to pay for a variety of goods and services (Lansing 2003)

2.2.12 Major Market Indicators

DJIA: Dow Jones industrial average which represents the 30 large publicly traded stocks in US. The value of the DOW Jones is the scaled average of the prices of its component stocks. This is mostly widely followed index which indicates the movement of the stock market. We use this index to study the influence of the macro economic variables on the stock markets. Since DJIA comprises the large industries in US, it is a good representation of the market as a whole.

S&P500: The SP500 Index is capitalization weighted index of prices of 500 large cap companies which are actively traded in US. After Dow Jones index, SP500 is the widely followed index. It acts as a leading indicator for the US economy. It comprises all kinds of industries, thereby representing the whole market. In this thesis, S&P500 is the most widely used index for almost all the analysis, including macro, sector and individual stock analysis. We forecast the SP500 Index to determine our investment strategy.

2.3 *Sector Analysis*

Identifying Best of Breed - Potential Market Leaders, followers, and natural hierarchical feeding chain (e.g. semiconductor manufacturers, equipment manufacturers, test equipment, etc). Sector analysis forms an important part of the framework, where the performance of individual sectors provides the information required for profitable investment. Sector performance contributes to almost 50-60% of our portfolio performance. Hence it is important to identify the right sectors during different time period. Individual sector performance varies depending upon the business cycle the economy is in.

In current thesis, ETF's which track individual sectors are used for analysis. All the companies which are traded are categorized into 10 sectors, which are again categorized into n different sub categories, based on the sector. Major sectors are Basic materials, conglomerates, industrial goods, consumer goods, healthcare, financial, technology, services, utilities, others. Some of the ETF's we use are SPDRs such as XLV,XLF,XLY etc. Apart from these few of the other sector tracking ETF’s are vanguard and iShare.

2.4 *Stock Analysis - Fundamentals*

Fundamental analysis is one of the important aspects of investment. It is a process of identifying a company, whose performance exceeds its peers and also the market as a whole and is expected to perform better in future, investing in which will result in greater profits. It involves careful reading of the companies, quarterly reports, annual statements and following it very closely. In this thesis, greater attention is given to fundamental analysis, since it helps in identifying the right stock to be a part of the portfolio.

2.4.1 Investments versus Trades

Investments and trade have lot of differences between them, even though they are aimed at making profits out of our investments. To explain in simple words, investments are for long term and less risky when compared to the trades which are short term with greater risks associated with them. When one wants to invest x amount of money and expects the returns after a definite period of time, without changing is positions on the portfolio then the investor is making an investment. However trading is lot more risky and requires deeper understanding of the markets dynamics and timing skills. Trading is done for shorter or very short duration where the individual buys in the morning and sells it in the evening. In this thesis, we focus on investments rather than trade, since investments are based on performance of the company, sector and the economy as a whole. Trading is mostly based on market movement, dependent on news and other shocks which are not trivial to model or for which we have sufficient data.

2.4.2 Profile, Company Web Site, Prospectus

One of the aspects of the fundamental analysis is to learn about the company, what core business is and what products they make and to analyze how good

the demand for such products in the market is. A generic knowledge on the company is very essential to understand its operations. As warren buffet says, when one buys a stock, they buy the part of the business.

2.4.3 Key Statistics

Key statistics refers to the fundamental ratios, which are used in fundamental analysis. These form the basis of the analysis. Understanding and obtaining the ratios are critical aspect. Some of the key statistics are, PE ratio, Price to sales ratio, Book value, Cash flow ratio, Market cap, Beta and etc. Looking at the key statistics gives us a good idea on the overall performance of the company and also if its a good investment. However before investing in any stock, one should perform a detailed analysis to make a decision.

2.4.4 Valuation

Valuation in fundamental analysis is aimed at provided a means by which one can identify the intrinsic value of a particular stock. In simple terms, it is a method to determine if we are paying the right value for the stock. Typically stocks are valued and classified into two categories, undervalued and overvalued. When a stock is said to be undervalued, it means that , the stock has a greater potential and the current market price does not reflect the future growth/potential of the stock. Hence learning this early, we can make profits before the stock reaches its expected intrinsic value. When a stock is said to be overvalued, it means that, the stock is trading at the higher price than the intrinsic value of the stock. Thereby going ahead the stock price is going to fall leading to a loss if invested at the current market price. There are many stock valuation tools available, while one of the most widely used tool/method is the discounted cash flow method or net present value. Apart

from this, valuation is also done based on the fundamentals of the company. In this thesis, we associate greater importance in analyzing the fundamental of the company.

2.4.5 Income Statement

Income statement is typically the profit or loss statement. It is self explanatory from the word that it contains all the details about the profits and loss of the company in a particular quarter or a fiscal year. Some of the key values given in the income statement are, Revenues, sales, Operating profit, Net income, total expenses and etc. Income statement helps an investor in analyzing the performance of a company in the past and also to predict the future performance by studying the statement in detail. In this thesis we use the income statement to compute and analyze some of the key fundamental ratios as mentioned above.

2.4.6 Balance Sheet

This indicates the firm's financial position at a fixed point of time. It contains details about what the firm's assets (fixed and current) are, what are its liabilities and what is the shares holders equity (Frank K. Reilly 2002). Some of the key values are, Current assets, current liabilities, share holders equity, long term and short term debt etc. In this thesis we use the balance sheet to compute and analyze some of the key fundamental rations such as debt ratio, operating profit margin and etc.

2.4.7 Cash Flow

Apart from the above two statements, the cash flow statements are a critical input for the analyst and the investors. It integrates the information on the balance sheet and income statement. This helps us determine the value of the

firm and also evaluate the risk and returns of the firm's stock (Frank K. Reilly 2002). In this thesis we use the cash flow to compute and analyze some of the key fundamental rations such as cash ratio, quick ratio, current ratio and etc.

2.4.8 Insider Trading and Holdings

By definition, insider trading involves purchase and sale of securities on the basis of important information that is not publicly available. This is typically carried out by the firm's managers or higher officials who have exclusive information in some of the key decisions of the company which would result in rise or fall of the stock price when made available to the public. SEC prohibits insider trading (SEC Rule 10b5-1). Even though access to such information would be of great use for investment, we don't focus on this in our current work.

2.4.9 Institutional Holdings

Institutions are mutual funds, pension plans, trust funds, and other large investors, and account for roughly 50 percent of all stockholdings. The presence of strong institutional sponsorship (large holdings) verifies that a stock is a viable growth candidate (lehman 2004)

2.4.10 Analyst Estimates

Analyst estimates are the company's future quarterly or annual earnings. They predict the performance of the company in the future; based on their past performance, balance sheets, income statements, cash flow statements, company policies, competitors, market share and etc. some of the common estimates are Earnings per share, Future sales, expected profits due to current or new product and etc. They also give recommendations on a

particular stock to buy, hold or sell. In recent times whisper number has been gaining importance among the investor community. It is the average of the earnings per share number given by number of analyst following a particular company. This number forms a basis of expectation for that quarter or year. This number is closely watched and markets reacts, by comparing the whisper number with the companies actual EPS.

2.5 *Technical Analysis Tools*

The fundamental idea behind Technical analysis is that, stock prices move in the trends that persist. This assumption of the technical analysis directly opposes the efficient market hypothesis (Frank K. Reilly 2002). Technicians believe that, the prices are stock move, based on particular information available to the market. However the striking point is, the information is not available to everyone at the same time. It takes time for the information to propagate, and hence by this, the person at the top of the information chain, can make profit, before it reaches everyone and the prices stabilize.

The impact of the news/information can cause the prices to go up or down and hence the amount of profit one can make by this is not abnormal. It also involves considerable amount of risk, since the movement is purely based on external events and mood of the overall market. It is greatly dependent on the past performance and trends of the stock and using that to predict the movement in the future. Below are the some of the most commonly used technical analysis tools.

2.5.1 Moving Average lines

Simple moving average lines are most commonly used tools. There are multiple moving average lines depending on the duration we are interested in

analyzing. Short term trend can be analyzed by seeing the 50 day moving average, and long term trend by the 200 day moving average. Each days MA value is the past n days average closing price/open price. This would smoothen out the high frequency data in the time series. These MA lines are seen with respect to the original price line.

If overall market trend is down, then the 50 Day MA line will be above the current price line with the similar trend and vice versa during an uptrend. When the current price line breaks / crosses the 50D MA line with considerable volume, that indicates the trend reversal and hence this point can be used as a buy point. The same analysis can be used with 200D MA line if we wanted to see long term trend. One other method to use them is comparing the 50D MA line and 200D MA line, where the crossing of the 50d MA line from below or above the 200D MA line would indicate a trend reversal. Averages can be computed in 2 ways. **Simple Moving Average** gives equal importance and weightage to all the values, where as **Exponential Moving Average** only the most recent prices are given more weightage than the older ones, hence giving us a good indicator of the trend in the movement of the stock.

2.5.2 Bollinger Bands

Famous technical trader developed this technique for trading. A band is plotted above and below the 21 day simple moving average line, representing the volatility with which the average prices are moving. The wider the band indicates more volatility and narrower the band indicates less volatility. The standard deviation is measure of volatility and hence the bands adjust themselves to the movement of the stock prices. The way it is used is, the

more the average moves towards the lower band, indicated the markets are oversold and hence we could see the reversal of trend and vice versa. (StockCharts n.d.)

Middle Band = 20-day simple moving average

Upper Band = 20-day SMA + (20-day standard deviation of price x 2)

Lower Band = 20-day SMA - (20-day standard deviation of price x 2)

2.5.3 Money Flow Index

Money flow index is similar to the RSI, however here volume is also taken into consideration where as RSI includes only the price. it is the measure of strength of money coming in and going out of a security and can be used to predict the reversal of the trend. It is computed by using the below formula (StockCharts n.d.)

- Typical Price = (High + Low + Close) / 3
- Money Flow = Typical price * Volume
- Money Ratio = Positive Money Flow/Negative Money Flow
- Money Flow Index = 100 - (100/ (1 + Money Ratio))

2.5.4 Parabolic SAR

It is created by the same person who came up with RSI. "*It is used to determine the direction of an asset's momentum and the point in time when this momentum has a higher-than-normal probability of switching directions* "- (Investopedia n.d.)

2.5.5 Relative Strength Index

RSI is a momentum oscillator that measures the speed and change of the price movements. It is an extremely popular momentum indicator (Frank K. Reilly 2002). This is the ratio of average of closing prices up days of particular

stock to the average of closing prices downs for the stock over a certain time period. This is usually 14 days (Investopedia n.d.). Relative strength is the ratio equal to the price of a stock relative to the value of some other stock or a market index such as SP500. This ratio is plotted over time and analyzed (StockCharts n.d.). If the ratio increases in trend, this indicated that the stock is performing better than the market and the trend is expected to continue, and when the ratio drops then it is underperforming than the market.

2.5.6 Slow and Fast Stochastic

Developed by George C. Lane in the late 1950s, the Stochastic Oscillator is a momentum indicator that shows the location of the close relative to the high-low range over a set number of periods. The Stochastic Oscillator "doesn't follow price, it doesn't follow volume or anything like that. It follows the speed or the momentum of price and hence can be used to identify bearish and bullish divergence to foreshow reversals (StockCharts n.d.)

%K = (Current Close - Lowest Low)/(Highest High - Lowest Low) * 100

%D = 3-day SMA of %K

Lowest Low = lowest low for the look-back period

Highest High = highest high for the look-back period

2.5.7 Volume

The basic idea behind this indicator is volume precedes price. It is a simple indicator where the volumes of the up periods are added and volumes of the down periods are subtracted. The cumulative total of these volumes gives us OBV line. This line is then used with the price chart to identify the buy and sell signals. For example, if today the closing price is greater than yesterday's closing price, then the new

OBV = Yesterday's OBV + Today's Volume

If today the closing price is less than yesterday's closing price, then the new

OBV = Yesterday's OBV - Today's Volume

If today the closing price is equal to yesterday's closing price, then the new

OBV = Yesterday's OBV

2.5.8 MACD- Moving average convergence divergence

This is an indicator which involves two other indicators, 26 day EMA and 12 day EMA. These two signals plotted with the current price chart, helps us identify the buy and sell spots.

2.5.9 Williams %R

William %R, sometimes referred to as %R, shows the relationship of the close relative to the high-low range over a set period of time. The nearer the close is to the top of the range, the nearer to zero (higher) the indicator will be. The nearer the close is to the bottom of the range, the nearer to -100 (lower) the indicator will be. If the close equals the high of the high-low range, then the indicator will show 0 (the highest reading). If the close equals the low of the high-low range, then the result will be -100 (the lowest reading). (StockCharts n.d.)

%R = [(highest high over n periods - close)/ (highest high over n periods - lowest low over n periods)] * -100

3 MACROECONOMIC ANALYSIS

3.1 *Overview*

In this chapter, Importance and influence of macroeconomic factors on the stock markets is presented. Time series forecasting models are used to predict the macro variables. Factor models are used to determine the influence of macroeconomic factor on stock markets. Fixed window and Moving window methods are used to improve the accuracy of prediction and are compared to determine the efficient method. Predicted data is used in sector analysis. Choosing different models, independent variables in the model, prediction methods, different window lengths and order of the models are few of the knobs available. Influence of variables such as housing, gold, markets on macro variables will be considered for future work.

3.1.1 Hypothesis

Macroeconomic variables such as GDP, money supply (m2), prime loan rate, unemployment rate, consumer spending, CPI, Treasury rate, consumer credit, industrial production is believed to have a significant effect on the stock market. These macro variables are used to predict the SP500 index returns n steps into the future.

3.2 *Modeling and prediction of macroeconomic variables*

3.2.1 Overview

Based on the historical data macroeconomic variables such as GDP are forecasted (n) steps into future for different time periods of economic expansion and contraction. Data is forecasted with two different methodologies, and each method implements two different models. Different methods are compared and analyzed to determine the forecast accuracy.

3.2.2 Modeling

Macroeconomic variables are predicted using an Autoregressive Moving average model.

3.2.2.1 Model 1: ARMA (p,q) : Auto regressive moving average.

ARMA (Box 1994) is a common model for forecasting time series data. The model includes AR (Auto regressive) terms which are lags of the series appearing in the equation and MA (Moving average) terms which are lags of the forecast errors.

'p' - represents the number of lags of the forecast series (AR terms)

'q' - represents the number of lags of the forecast errors in prediction equation

The generic ARMA model is represented by

$$r_t = \emptyset_0 + \sum_{i=1}^{p} \emptyset_i r_{t-i} + a_t - \sum_{i=1}^{q} \theta_i a_{t-i} \qquad (3.1)$$

Where a_t is white noise series and p and q are non-negative integers. Time series data is pre-processed. Forecast a stationary series which has constant statistical properties like mean, variance and constant autocorrelations over a period of time leads to accurate predictions for that time period. Most of the time series data is assumed to be stationary, and is achieved by using mathematical transformations such as differencing.

Differencing is a technique where each value of the time series is the result of the difference between its previous value and its current value. Mostly this step will result in a stationary series. However need arises to double difference based on the data under study. The model is represented as

Auto regressive Integrated Moving average -ARIMA (p,d,q) having similar model structure as ARMA but with an additional term representing the order of differencing (d).

3.2.2.2 Identifying the order of differencing:

Time series data is differenced before being fit to the model. Data can be over differenced or under differenced. The order of differencing is determined by autocorrelation plot. The right amount of differencing is the lowest order which results in a series which has near constant mean and has the auto correlation function plot decaying exponentially to zero, either from above or below. Slower decay indicates a significant auto correlation till higher lags, and hence the data is differenced until an exponential decay is obtained. Differencing of series is stopped when the autocorrelation with lag-1 is 0 or negative. Further differencing of data will cause the autocorrelation with lag -1 to more negative where the problem of over differencing arise.

3.2.2.3 Identifying the order of AR and MA terms:

Order of the Auto Regressive and Moving average terms can be determined by many methods. Common methods to identify them are ACF and PACF plots.

Auto correlation is the correlation between members of series of observations ordered in time (Buckland 1971). It is the degree of similarity between a given time series and a lagged version of the same time series over successive time intervals. (Investopedia n.d.)

ACF: Auto correlation function

ACF at lag k is denoted byρ_k, is defined as

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{3.2}$$

Where γ_k= covariance at lag k and γ_0 = variance

Auto Correlation function helps us determine the number of MA terms.

PACF: Partial auto correlation function

It is measure of correlation between the observations that are k time periods apart after controlling for correlation at intermediate lags. In other words partial autocorrelation is the correlation between Y(t) and Y(t-k) after removing the effect of the intermediate Y's. Partial auto correlation removes the influence of the intervening variables. (Damodar N Gujarati 2007). PACF plot helps us determine the number of AR terms required.

3.2.2.4 Model 2: ARIMA (p,d,q) with Regressors

Macro variables such as consumer spending and money supply (Damodar N Gujarati 2007) influence GDP value. Lagged GDP value also influences the current GDP value along with other macro variables. This leads to a Hybrid model where independent variables/regressors are included along with the ARIMA model. Generic representation of the model is given in equation (3.3)

$$r_t = \emptyset_0 + \sum_{i=1}^{p} \emptyset_i r_{t-i} + a_t - \sum_{i=1}^{q} \theta_i \, a_{t-i} + \sum_{i=1}^{n} \varphi_i \, x_i \tag{3.3}$$

Note: In this model the AR part apply to both the regressors and the series of interest.

3.2.3 Data

Data is obtained from NBER (National bureau of economic research) and Federal Reserve Bank of St Louis which release the macro economic data

periodically. GDP (Billions of dollars) - Seasonally adjusted Annual rate released quarterly.

Date: From 1947 Q1 to 2010 Q1 for MODEL 1

Date: From 1959 Q1 to 2009 Q3 for MODEL 2

Consumer Spending (billions of dollars) and Money supply (M2) (Billions of dollars)

Date: 1959 Q1 to 2009 Q3 for MODEL 2

3.2.4 Data Preprocessing

GDP quarterly data is transformed into a returns series to fit to ARMA model. The quarterly data is annualized by multiplying by 4 in order to compare the annual growth. ACF plot indicates the need for differencing the GDP time series. Log of the GDP series is differenced to obtain the return series as difference of the log series is equivalent to computing the return between current and previous quarter. Key indicators of GDP are determined by correlation coefficients of GDP quarterly series and other macro variable series. Out of many consumers spending and money supply have high correlation and significant pValue with GDP. Hence there macro variables are used to forecast the GDP into future.

3.2.5 Implementation

GDP data is forecasted n steps into future using two different methodologies. Each method implements 2 models. The lags of the models are determined as AR=1 and MA =2 from ACF and PACF plots. Hence these are used for model fit and prediction.

Model 1: ARIMA (1, 1, 2)

Forecast GDP using past GDP values and white noise.

$$y_t = \emptyset_0 + \sum_{i=1}^{p} \emptyset_i y_{t-i} + \sum_{i=1}^{q} \theta_i \, a_{t-i} \qquad (3.4)$$

Where

y_t = GDP value one step into future.

a_t = white noise term

$\emptyset_0, \theta_i$ = Coefficients of their corresponding AR or MA terms

Model 2: ARIMA (1, 1, 2) with Regressors

Forecast GDP using past GDP values, Independent/explanatory variables and white noise.

$$y_t = \emptyset_0 + \sum_{i=1}^{p} \emptyset_i y_{t-i} + \sum_{i=1}^{q} \theta_i \, a_{t-i} + \sum_{i=1}^{n} \varphi_i \, x_{t-i} \qquad (3.5)$$

Where,

y_t = GDP value one step into future.

a_t = white noise term

x_t = Independent variables

$\emptyset_0, \theta_i, \varphi_i$ = Coefficients of AR, MA and Regressor terms respectively

3.2.6 Methodology

1. Fixed Window Prediction
2. Moving Window Prediction

In the above mentioned methodologies, the entire data is split into Training data and validation data. **Training data** is used to fit the model and estimate

the coefficients and **Validation data** is used for validation. Forecasted/estimated data is compared with the validation data to determine the accuracy of prediction. These methodologies are employed for various time intervals, to identify the patterns during different time periods of economic contraction and expansion.

3.2.7 Forecasting and Analysis

GDP is forecasted for different time periods, categorized based on the economic conditions.

- Boom Period: This is period where the economic expansion or growth in the economy.
- Bust period: This is the period of economic contraction where the economy grows in the negative direction also called as recession.

From the past knowledge, 2000-2002 and 2008-2010 were the most recent recessions, 2004-2007 and 1996-2000 were considered as growth period where in the economy was expanding. Model is fit to the training data in each time period to estimate the coefficients. The estimated coefficients are used to forecast the GDP data.

- Boom period: 1993(Q1)-1999(Q4), 2004(Q1) -2007(Q4)
- Bust period : 2000(Q1)-2003(Q4), 2008(Q1)-2010(Q1)
- Boom and bust period: 1996(Q1)-2003(Q4)
- Bust and boom period: 2000(Q1)-2007(Q4)

The Forecasted GDP series for the below periods are compared with the real GDP series to analyze the accuracy of prediction.

- Boom period: 1993(Q1)-1999(Q4), 2004(Q1) -2007(Q4)
- Bust period : 2000(Q1)-2003(Q4), 2008(Q1)-2010(Q1)

3.2.7.1 Method 1: Fixed Window Training and Prediction Method

In this method, the below steps are performed.

1. Model 1 is fit to the chosen training data.
2. The coefficients or weights are estimated by least square method.
3. The estimated coefficients are used to forecast the GDP series n steps into future.
4. N steps correspond to the chosen validation period.
5. Mean, maximum and minimum error percentages between Predicted data and the real data Worst case error are computed.
6. The same procedure is repeated using model 2.

The forecasted data is one step ahead, from the current available data, ie, GDP data of next month is predicted with the current month's available data.

3.2.7.2 Method 2: Moving Window training and prediction method

In this method, the oldest data is neglected and latest data obtained is used to forecast the next quarter data. By doing this, the prediction is more accurate since the economic activity is continuously being tracked whenever we include the new data obtained. Below are the steps followed to implement this method.

1. A fixed window length of l is chosen
2. The data from the training period of length l is used to fit to the model 1
3. Coefficients/ weights are estimated from the fit.
4. These estimated coefficients are used to forecast / predict n steps into future.

5. The window is moved, i.e., first data is neglected and new data point is included to fit the model to the training data set.
6. Coefficients are estimated and these are used to forecast the next data.
7. This operation is repeated till all the data for the above required Prediction time period is forecasted.
8. The window length is changed as per our requirement and steps 1 to step 7 are repeated.

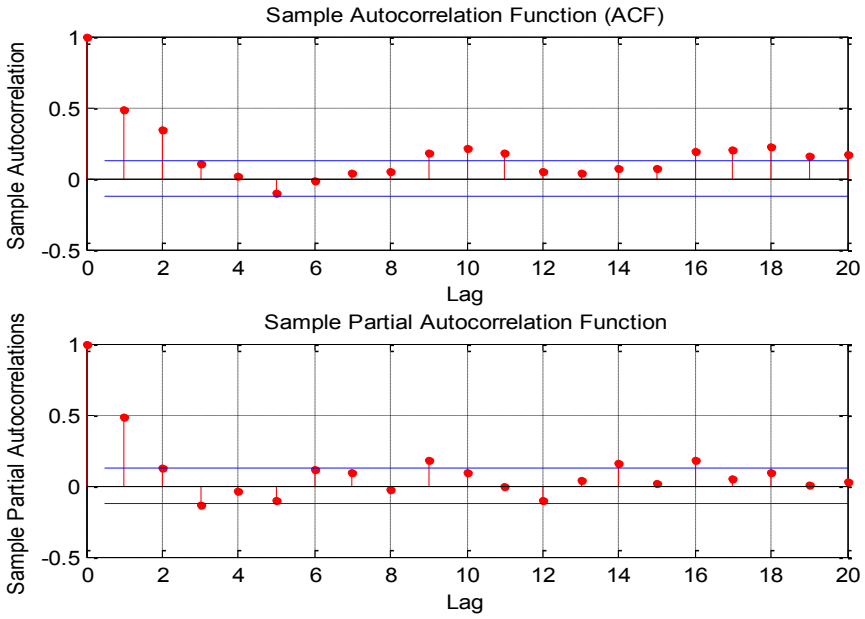

Fig 2: Auto Correlation and Partial Auto Correlation Plots

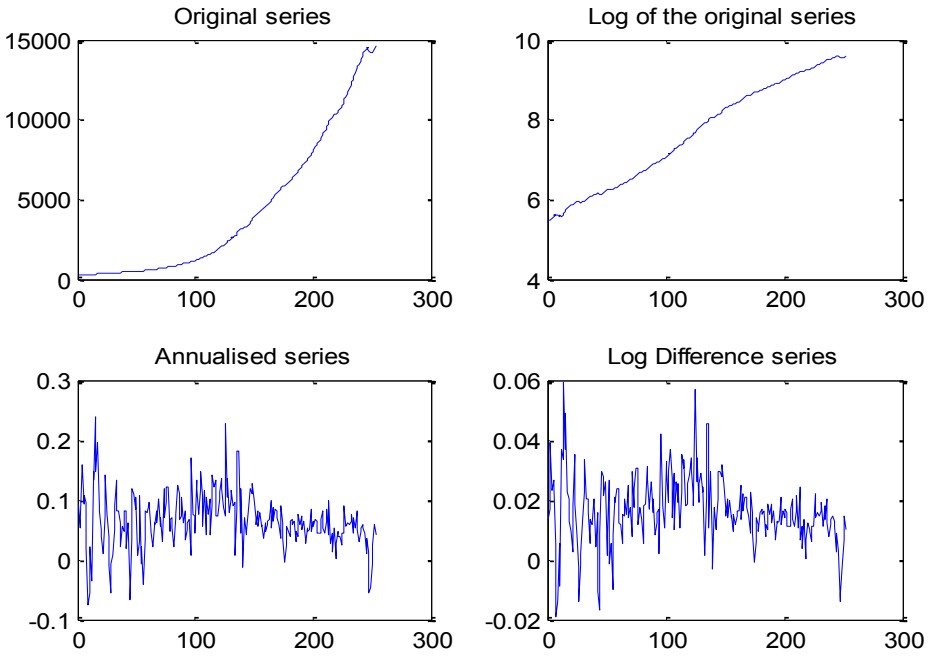

Fig 3: Time Series Plots of Original GDP Data, Log of GDP data, Log Difference of GDP data

3.2.8 Results

Table 1 shows the percentage error between the Real GDP values and Predicted GDP value using fixed window and moving window methods for Model 1 and Model 2.

Model 1: ARIMA (1, 1, 2)

Model 2: ARIMA (1, 1, 2) - with Independent variable [Consumer Spending and Money Supply]

Table 1: Percentage Error btw real and predicted GDP by Fixed Window method

Train Period	Forecast Period	Model 1	Model 2
93-99 (UP)	2000-2002 (Down)		
	Mean Error	3.52%	2.24%
	Min Error	1.07%	0.02%
	Max Error	8.47%	4.58%
2000-2003 (Down)	2004-2007 (UP)		
	Mean Error	4.65%	1.16%
	Min Error	0.66%	0.05%
	Max Error	7.43%	3.16%
2004-2007 (UP)	2008-2010 (Down)		
	Mean Error	7.97%	57.72%
	Min Error	1.32%	0.51%
	Max Error	13.10%	100%
96-2003 (UP+Down)	2004-2007 (UP)		
	Mean Error	1.96%	1.79%
	Min Error	0.30%	0.11%
	Max Error	2.98%	3.30%
2000-2007 (Down+ UP)			
	2008-2010 (Down)		
	Mean Error	249.48%	9.21%
	Min Error	24.57%	3.01%
	Max Error	654.75%	14.75%
1993-1999 (UP)	2003-2007 (UP)		
	Mean Error	5.51%	0.80%
	Min Error	1.48%	0.03%
	Max Error	8.96%	1.97%

2000-2003 (Down)	2008-2010 (Down)		
	Mean Error	4.62%	3.52%
	Min Error	0.68%	0.56%
	Max Error	7.60%	8.51%

Table 2: Percentage Error btw real and predicted GDP by Moving Window method

Time window	For M1 At wLength (No of Win Len tried)	Model 1	Model 2 Error @ M1 wLengths	For M2 At wLength (No of Win Len tried)	Model 2	Model 1 Error @ M2 wLengths
93-99 (UP)						
Min WCE (%)	25 (27)	1.61	5.83	24 (27)	5.83	1.61
Max WCE (%)	5 (27)	10.31	8.9	9 (27)	9.59	2.01
2000-2003 (Down)						
Min WCE (%)	6 (15)	1.69	4.92	6 (15)	4.92	1.69
Max WCE (%)	7 (15)	5.88	4.92	8 (15)	5.70	2.20
2004-2007 (UP)						
Min WCE (%)	15 (15)	1.29	5.12	9 (15)	5.12	12.85
Max WCE (%)	9 (15)	12.85	5.12	7 (15)	9.01	1.90
1996- 2003 (UP + Down)						
Min WCE (%)	30 (31)	1.13	4.91	18 (31)	4.91	7.20
Max WCE (%)	12 (31)	11.94	7.14	13 (31)	9.15	7.70
2000-2007 (Down+ UP)						
Min WCE (%)	31 (31)	1.29	5.02	28 (31)	5.02	1.50
Max WCE(%)	15 (31)	18.31	6.50	8 (31)	8.07	6.50

WCE: Worst Case Error

The below figures are the best predictions among all the time periods using moving window and fixed window methods with both the models.

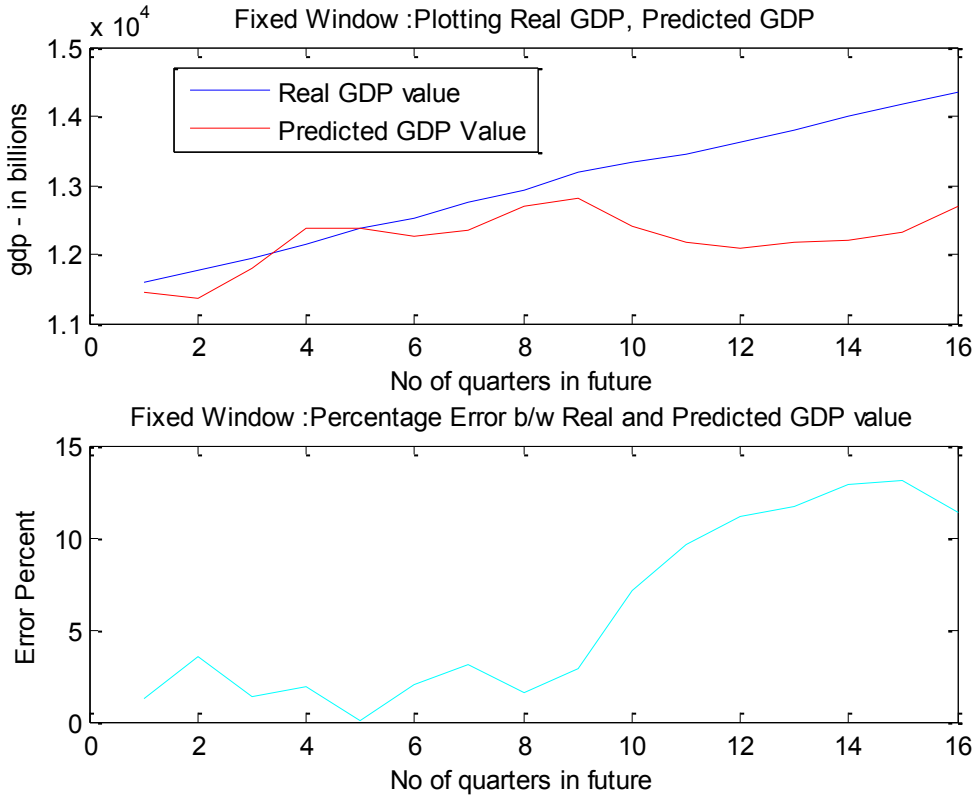

Fig 4: GDP Prediction using fixed window method using model 1 for time period 2003-2007

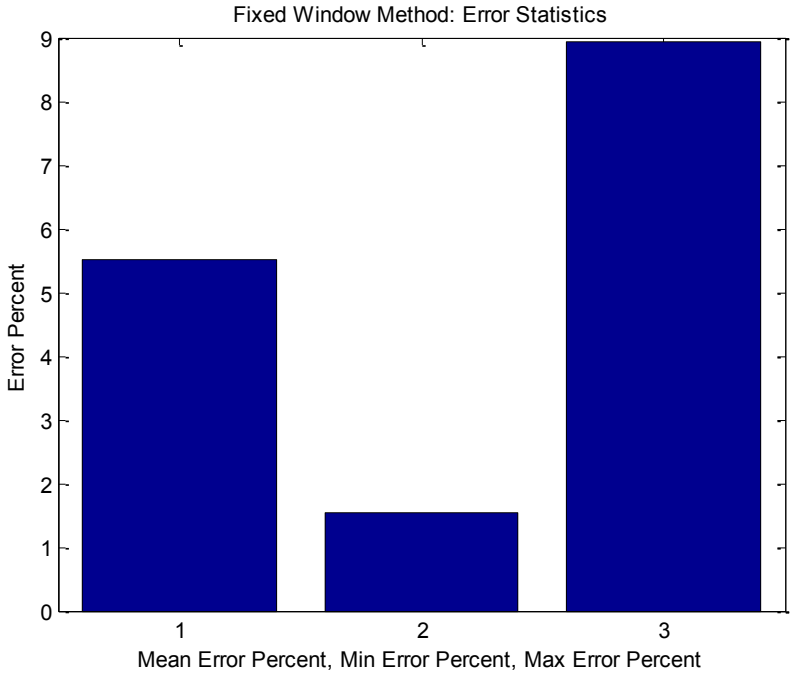

Fig 5**:** Mean, minimum and maximum error between real and predicted GDP using fixed window method with model 1 for time period 2003-2007

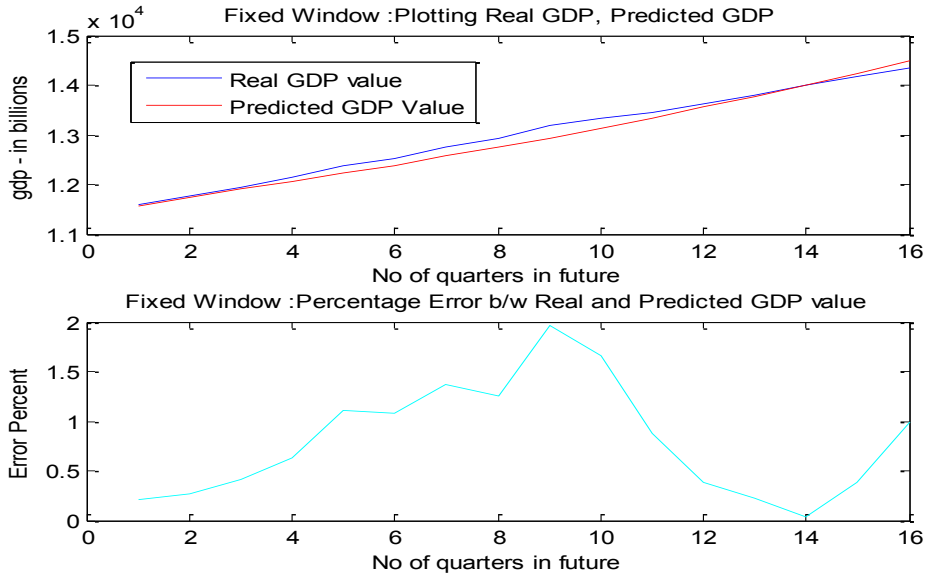

Fig 6: GDP Prediction using fixed window method using model 2 for time period 2003-2007

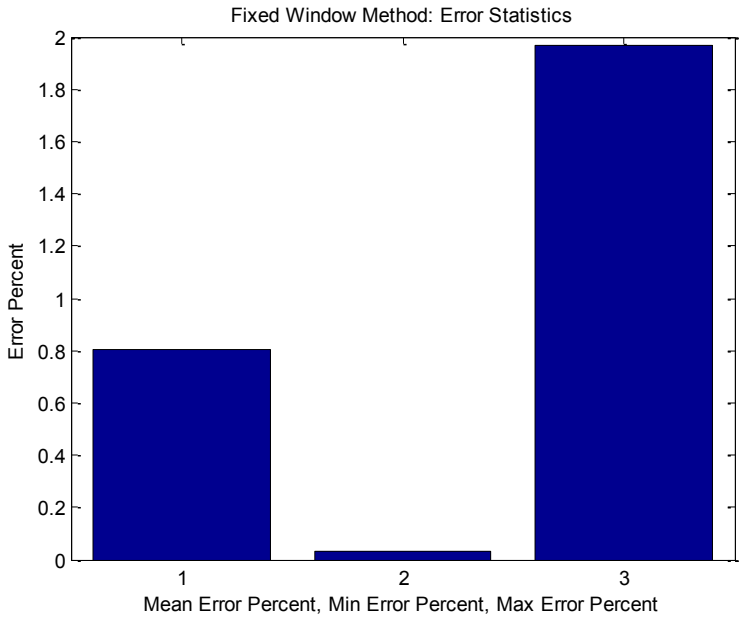

Fig 7: Mean, minimum and maximum error between real and predicted GDP using fixed window method with model 2 for time period 2003-2007

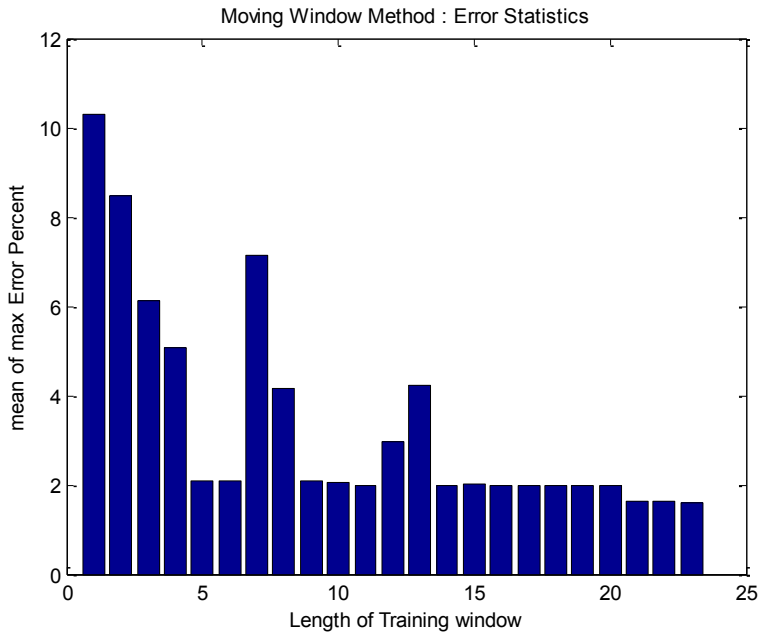

Fig 8: Worst case error between real and predicted GDP values using Moving Window method with Model 1 at different window lengths

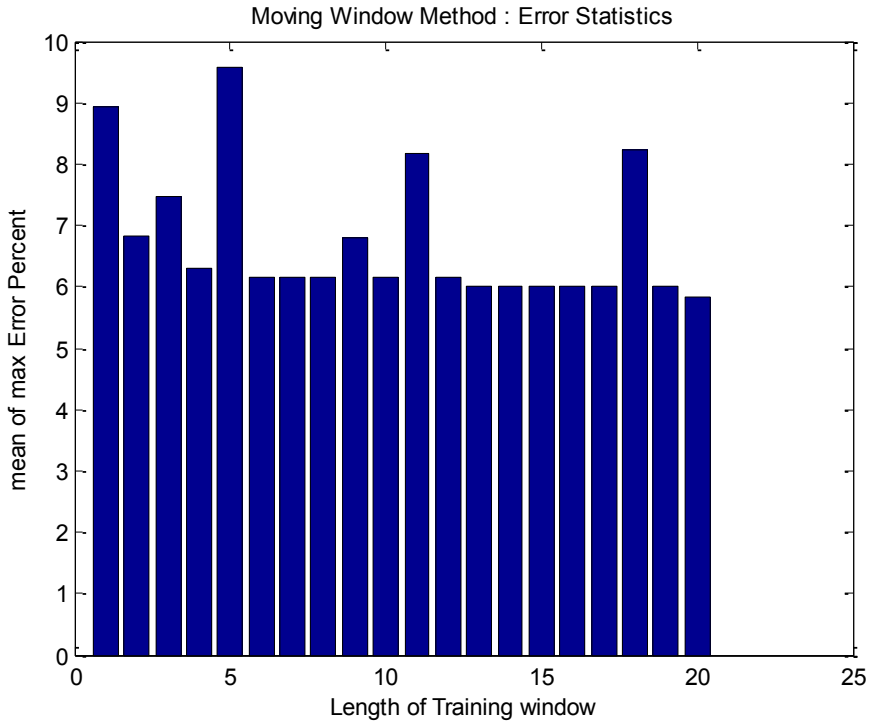

Fig 9: Worst case error between real and predicted GDP values using Moving Window method with Model 2 at different window lengths

3.2.9 Summary

In Table 1

1. For the Fixed Window method, Model 2 (Fig 7) has LESS error than Model 1 (Fig 5)
2. Consumer spending and Money supply are key variables which predict GDP more accurately.
3. Error decreases when we train for Boom /Bust and predict for Boom /Bust respectively
4. Larger training window including boom and bust predicts more accurately than smaller training window

In Table 2,

1. There is no Considerable difference between Model 1 and Model 2.
2. Prediction based on latest information of GDP every quarter is more accurate.
3. No need for Independent variables.
4. Greater the window length, better the prediction (Fig 8)

3.3 *Impact of macroeconomic variables on markets*

3.3.1 Overview

In this section the impact of macroeconomic variables on stock markets is analyzed. To forecast the market (SP500) movement with the knowledge of macroeconomic variables, forecasted data from section 3.4 is used.

3.3.2 Modeling

SP500 index (defined in chapter 2) is chosen as the representative of the overall market. It is believed that the movement in markets is due to the change in the economic conditions. Hence macro economic variables are included as a part of the model to forecast the SP500 index.

3.3.2.1 Model 1: Multiple linear regressions

In this model, two or more independent explanatory variables and the response variable are included and relationship between them is established. (Damodar N Gujarati 2007)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} \qquad (3.6)$$

Where

Y_i = Month ahead Log S&P Value - (Response variable)

X_{2i}, X_{3i} = Macro Variables like GDP, Money Supply, Consumer Spending, Unemployment etc (explanatory variables)

β_2, β_3 = Coefficients of Macro Variables.

β_1 = Regression Constant.

3.3.2.2 Model 2: Hybrid model

In this model, a relation between the dependent variable and independent variables is established. However the independent variable not only depends on macro economic variables, but also an auto regressive term.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \sum_{i=1}^{p} \emptyset_{\mathrm{i}} \mathrm{y}_{\mathrm{t-i}} \qquad (3.7)$$

Where,

Y_i = Month ahead Log S&P Value - (Response variable)

X_{2i}, X_{3i} = Macro Variables like GDP, Money Supply, Consumer Spending, Unemployment etc (explanatory variables)

β_2, β_3 = Coefficients of Macro Variables.

β_1 = Regression Constant.

3.3.2.3 Model 3: Polynomial regression model

In this model, a relationship is established between the dependent variable and squares of the independent variables. The highest power of the independent variable is the degree of the polynomial. (Damodar N Gujarati 2007)

$$Y_i = \beta_1 + \beta_2 X_2^2 + \beta_3 X_3^2 \quad (3.8)$$

Where

Y_i = Month ahead Log S&P Value - (Response variable)

X_{2i}, X_{3i} = Macro Variables like GDP, Money Supply, Consumer Spending, Unemployment etc (explanatory variables)

β_2, β_3 = Coefficients of Macro Variables.

β_1 = Regression Constant.

3.3.3 Data

Data is obtained from NBER website and Federal Reserve Bank of St Louis which release the macro economic data periodically. GDP (Billions of dollars) - Seasonally adjusted Annual rate released quarterly. Consumer Price (Index), 10 Year Treasury (Rate), Prime loan (rate), Outstanding Cons Credit (Billions), Industrial Prod(Index), Money Supply(Billions), Unemployment (rate) and SP500 index value.

Date: From 1959 M1 to 2010 M1 for MODEL 1

Date: From 1959 M1 to 2010 M1 for MODEL 2

PACF plot with a higher Partial auto correlation at lag 1 gives the number of AR terms required.

3.3.4 Data Preprocessing

The available data is pre processed as per the requirements. Quarterly GDP data is converted to monthly data, since all the other macro economic data is obtained monthly. Linear data interpolation is done to obtain the monthly GDP data.

1) Log of SP500 index value is used for estimation and forecasting using the Models. There is no significant correlation between the SP500 Month ahead or past returns and the macro variables.
2) Among the available macro variables, prime loan rate is totally insignificant in predicting the SP500 value. Hence Prime loan rate is removed from the analysis as there is no significant correlation between the two variables.

Below is correlation analysis table showing the correlation between log (SP500 Index) and macro economic variables.

Table 3: Correlation Coefficients btw macro variables and log S&P500 index/ S&P500 Month Ahead returns

Macro Variables	Correlation With log S&P Index value (Month Ahead)			Correlation With SP return (Month Ahead)	
	Corr Coef	**P value**		**Corr Coef**	**P value**
CPI	0.9692	0		0.0053	0.8958
10 Year Treasury Rate	-0.2468	0		0.0277	0.4944
PrimeLoan	-0.0766	0.0581		0.0166	0.6822
Consumer Credit	0.9381	0		-0.0242	0.5491
Industrial Production	0.9756	0		-0.0162	0.6885
Money Supply	0.939	0		-0.0148	0.7153
Unemployment	-0.1277	0.0015		0.0963	0.0171
GDP	0.9619	0		-0.0134	0.7397

3.3.5 Methodologies

1. Update Window Prediction
2. Moving Window Prediction

In the above mentioned methodologies, Recursive least square and Kalman Filters tools are used and the entire data is split into Training data and validation data.

Training data is used to fit the model and estimate the coefficients and **Validation data** is used for validation. Forecasted/estimated data is compared with the validation data to determine the accuracy of prediction. These methodologies are employed for various time intervals, to identify the patterns during different time periods of economic contraction and expansion.

3.3.6 Implementation

Training Time periods:

SP500 is forecasted for different time periods, categorized based on the economic conditions.

- Boom Period: This is period where the economic expansion or growth in the economy.
- Bust period: This is the period of economic contraction where the economy grows in the negative direction also called as recession.

From the past knowledge, 2000-2002 and 2008-2010 were the most recent recessions, 2004-2007 and 1996-2000 were considered as growth period where in the economy was expanding. Model is fit to the training data in each time period to estimate the coefficients.

1. Boom period: 1993(Q1)-1999(Q4), 2004(Q1) -2007(Q4)
2. Bust period: 2000(Q1)-2003(Q4), 2008(Q1)-2010(Q1)
3. Boom and bust period: 1996(Q1)-2003(Q4)
4. Bust and boom period: 2000(Q1)-2007(Q4)

The Forecasted values for the below periods are compared with the real log (SP 500) value to analyze the accuracy of prediction

1. Boom period: 1993(Q1)-1999(Q4), 2004(Q1) -2007(Q4)
2. Bust period : 2000(Q1)-2003(Q4), 2008(Q1)-2010(Q1)

Method 1: Update Window Training and Prediction Method

The prediction is more accurate by incorporating the latest data to the existing training data. Coefficients which are estimated are constantly updated. This is very similar to the state estimation method for which kalman filter is widely used. The coefficients represent the states which are updated as and when new observation is obtained. The below steps are performed in the method.

1. Model 1 is fit to the training data for a particular period.
2. Initial window of length l is chosen to fit the model and estimate the coefficients.
3. The coefficients or weights are estimated by recursive least square method.
4. The estimated coefficients are used to forecast the GDP series n steps into future.
5. N steps corresponds to the number of months ahead we want to predict
6. Update the new data to the intial window of length L , thereby increasing the window length.
7. Steps 2 to 6 are repeated.
8. Worst case error(%) between the real SP500 and the predicted value is computed for the entire period of prediction.

9. The same procedure is repeated using model 2.

The forecasted data is one step ahead ie, (next month SP500 value), from the current date. Similarly the entire procedure is repeated for 2 month ahead, 3 month ahead and 4 month ahead prediction.

Method 2: Moving Window training and prediction method

In this method, the oldest data is neglected and latest data obtained is used to forecast n steps into future. This makes sure that the recent economic activity is given more importance rather than the entire past history.

1. A fixed window length of l is chosen
2. The data from the training period of length l is used to fit to the model 1
3. Coefficients/ weights are estimated from the fit.
4. These estimated coefficients are used to forecast / predict n steps into future.
5. The window is moved, ie, first data is neglected and new data point is included to fit the model to the training data set.
6. Coefficients are estimated and these are used to forecast the next data.
7. This operation is repeated till all the data for the above required Prediction time period is forecasted.
8. The window length is changed as per our requirement and steps 1 to step 7 are repeated.

3.3.7 Results

Using the above methods, 4 months ahead Log SP 500 values were predicted and the worst case error in each scenario was computed. Worst case error is

defined as maximum error between the predicted SP500 value and real values. The worst case error is computed among

a) n step predictions at any instance with the fixed window length
b) worst case error for the entire period for a fixed window length
c) Worst case error for the entire period among all the window lengths.

Table 4: Percentage Error btw real and predicted log S&P500 value for N months ahead in different time periods using update window method

Update Window Method	**1 Month Ahead Prediction**		**2 Month Ahead Prediction**		**3 Month Ahead Prediction**		**4 Month Ahead Prediction**	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Time Window	max Error (%)	max Error (%)	max Error (%)	max Error (%)	max Error (%)	max Error (%)	max Error (%)	max Error (%)
93-99	2.89	3.0391	3.085	3.124	3.366	3.293	3.189	3.127
2000-2002	3.67	3.8578	2.482	2.495	4.037	4.163	3.812	4.023
2003-2007	4.47	4.5445	6.115	6.541	2.883	2.944	2.753	3.083
2008-2010	3.11	3.059	4.297	3.799	4.284	3.855	4.053	4.643
1996-2003	3.99	4.0743	4.008	3.683	4.316	3.604	3.925	3.593
2000-2007	3.67	3.8578	2.482	2.495	4.037	4.163	3.812	4.023
1993-2010	5.928	4.6968	6.977	5.886	7.785	6.741	8.343	7.391

Table 5: Percentage Error btw real and predicted log S&P500 value for N months ahead at different window length in different time periods using moving window method

Moving Window Method		**1 Month Ahead Prediction**					**2 Month Ahead Prediction**			
		Model 1		Model 2			Model 1		Model 2	
Time Window	*Worst Case Error(WCE)*	*Error (%)*	*wLen*	*Error (%)*	*wLen*		*Error (%)*	*wLen*	*Error (%)*	*wLen*
93-99	min WCE	2.39	13	2.737	20		2.67	19	2.82	20
	max WCE	35.76	63	29.92	63		22.44	66	33.60	59
	Mean WCE	7.324		7.593	NaN		7.61		9.84	NaN
2000-2002	min WCE	1.97	24	2.189	34		2.47	20	2.32	35
	max WCE	5.43	31	6.507	31		10.98	9	8.59	32
	Mean WCE	2.97		3.731	NaN		3.83		4.39	NaN
2003-2007	min WCE	2.06	31	2.205	31		1.67	52	1.717	54
	max WCE	19.69	56	23.21	58		15.95	60	17.12	60
	Mean WCE	8.36		8.606	NaN		5.40		5.35	NaN
2008-2010	min WCE	2.39	14	2.594	13		3.30	16	3.34	16
	max WCE	6.37	9	6.158	9		8.13	9	6.6	9
	Mean WCE	3.46		3.447	NaN		4.02		4.10	NaN
1996-2003	min WCE	3.02	13	3.185	13		2.90	44	2.93	44
	max WCE	26.17	31	32.97	30		25.37	31	32.04	31
	Mean WCE	12.13		13.83	NaN		9.77		11.43	NaN

Moving Window Method		**3 Month Ahead Prediction**					**4 Month Ahead Prediction**			
		Model 1		Model 2			Model 1		Model 2	
Time Window	*Worst Case Error(WCE)*	*Error*	*wLen*	*Error*	*wLen*		*Error*	*wLen*	*Error*	*wLen*
2000 - 2007	min WCE	1.97	24	2.163	39		2.47	20	2.36	38
	max WCE	56.39	78	78.66	78		57.37	79	80.61	79
	Mean WCE	14.31		16.44	NaN		14.17		16.06	NaN
1993 - 2001	min WCE	2.88	7	3.162	20		3.48	12	3.73	21
	max WCE	105.36	183	540.9	189		131.3	183	675.5	189
	Mean WCE	14.79		51.48	NaN		18.11		70.98	NaN
1993 - 1999	min WCE	2.84	38	2.89	38		2.67	39	2.75	39
	max WCE	25.27	59	35.2	59		21.59	65	27.54	81
	Mean WCE	9.16		11.10	NaN		9.50		12.20	NaN
2000 - 2002	min WCE	2.24	20	2.26	20		2.48	36	2.57	36
	max WCE	8.40	31	9.73	33		7.50	32	8.66	32
	Mean WCE	4.13		4.76	NaN		4.17		4.53	NaN
2003 - 2007	min WCE	1.60	53	1.6	55		1.74	55	1.81	26

Moving Window Method		**3 Month Ahead Prediction**					**4 Month Ahead Prediction**			
		Model 1		Model 2			Model 1		Model 2	
Time Window	*Worst Case Error(WCE)*	*Error*	*wLen*	*Error*	*wLen*		*Error*	*wLen*	*Error*	*wLen*
	max WCE	17.44	60	19.95	60		14.96	60	14.69	60
	Mean WCE	5.20		4.965	NaN		5.85		5.249	NaN
2008 - 2010	min WCE	3.78	22	4.02	12		3.97	9	4.29	9
	max WCE	14.07	9	14.14	9		7.926	13	8.08	13
	Mean WCE	5.14		5.10	NaN		5.20		5.88	NaN
1996 - 2003	min WCE	2.72	42	2.79	42		2.69	40	2.74	40
	max WCE	23.19	34	29.2	32		19.18	33	23.20	33
	Mean WCE	7.96		10.15	NaN		7.38		9.52	NaN
2000 - 2007	min WCE	2.40	19	2.40	40		2.48	36	2.42	38
	max WCE	54.39	80	77.0	80		49.37	81	67.36	81
	Mean WCE	12.52		14.3	NaN		13.07		14.22	NaN
1993 - 20010	min WCE	3.78	14	4.06	20		3.52	13	3.53	109
	max WCE	118.90	183	602.	193		52.93	184	340.3	203
	Mean WCE	18.16		65.2	NaN		13.22		43.20	NaN

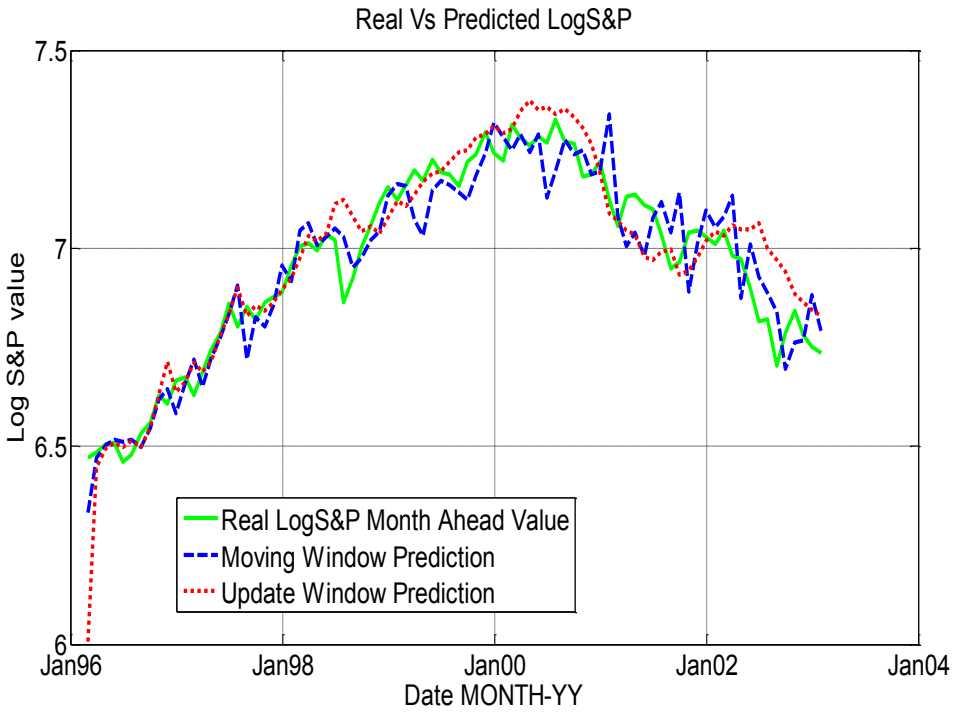

Fig 10: Log S&P prediction using Model 1 in 1996-2002

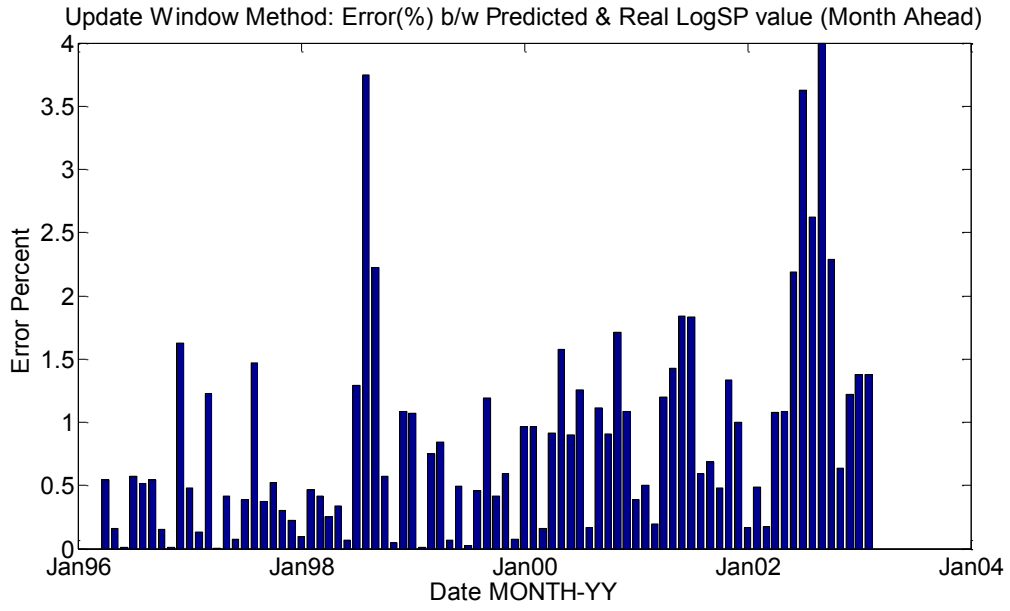

Fig 11: Error (%) between real and predicted Log SP 500 value using update window method for time period 1996-2002

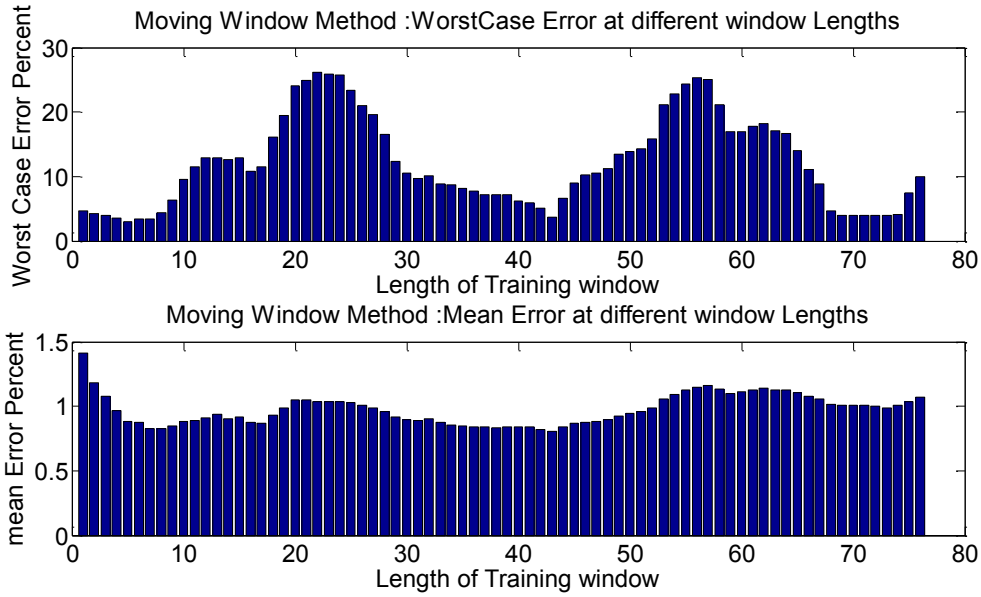

Fig 12: Max and Mean error (%) between real and predicted Log SP 500 value using moving window method with different window lengths for time period 1996-2002

3.4 *Observations*

1. Linear Regression Model gives more accurate Prediction. Independent variables are Macro Variables such as GDP, Unemployment, Money Supply, Interest Rate, Industrial production, Outstanding consumer credit and CPI
2. Non Linear Regression Model – No significant Improvement compared to linear model where Macro variables are quadratic.
3. Linear Regression with S&P value + macro Variable as Independent variables results is not encouraging. Sometime leads to greater error.

4. Macro variables have greater correlation coefficients with the S&P 500 index value.
5. No or very less significant correlation between Macro Variables and S&P returns.
6. Prime Loan has less correlation hence not considered for regression.
7. Update window method gives better prediction than Moving window.
8. Moving window method leads to greater errors at certain window lengths.
9. Lesser Window Lengths in Moving Window Method (for training) gives better prediction.

4 SECTOR ANALYSIS

4.1 *Introduction*

In this chapter, importance of sector analysis is explored in detail. State space asset model is used to forecast asset returns. Sector ETFs representing specific industries are considered as assets. Seven important macro economic variables GDP, MS, Unemployment rate, Ten-Year Treasury rate, Outstanding Consumer Credit, Consumer Price Index, and Industrial Production Index are chosen for the factor model. Recursive least square and exponential weighted least square methods are explored to improve the prediction. Model structure, independent variables in the model, prediction methods, training window length for prediction and frequency of returns (daily or monthly returns) can be changed. Influence of interest rates, inflation, currency, and commodities on sector performance will be considered for future analysis.

4.2 *Modeling and prediction of sector performance*

4.2.1 Modeling

Sector performance is based on the macro variables properties and economic activity; hence macro variables are used to predict the sector returns.

Model

$$\bar{y}^{k+1} = M^k \bar{y}^k + \bar{e}^k \qquad (4.1)$$

Where

- $\bar{y}^k$: $[\bar{r}^T \ \ \bar{x}^T \ ]^T$ is the asset return model's state vector for period k
- $\bar{r}^T$: Denotes the asset return vector
- M^k : (Estimated) state matrix for period k

- $\bar{e}^k$: Noise
- $\bar{x}^T$: Macroeconomic data for period k

State matrix is estimated for every new data obtained using the recursive least square technique. It is of the form

$$M^k = \begin{bmatrix} M_{11}^k & M_{12}^k \\ 0 & M_{22}^k \end{bmatrix} \quad (4.2)$$

Where

- M_{11}^k : Asset returns submatrix for period k
- M_{12}^k : macroeconomic factor to asset coupling submatrix for period k,
- M_{22}^k : macroeconomic submatrix for period k

4.2.2 Data

Three different sector ETF's are considered. Each ETF have individual ETF's tracking individual sectors. SPDR - Start from Dec 1998, iShare - from Nov 2000 and Van - from Aug 2001. The sector ETF's are downloaded from finance.yahoo.com.

4.2.3 Data Preprocessing

Macro data such as GDP, Consumer Index, Money Supply and etc increase exponentially. To fit the data to linear regression model, data is preprocessed to induce weak stationary by the process of differencing discussed in (Tsay 2005). Interest rate and unemployment rate are not modified as the data is in the required form.

4.2.4 Implementation

Month ahead average daily Sector returns and monthly returns are predicted using the macro variables. Available data between Dec 1998 and Jan 2010 are split into training period and validation period. The length of the training window changes based on the ETF's being predicted. The validation window starts from Jan 2005 for all 3 ETF's under study. Each of the macroeconomic series is forecasted using the auto regressive model. Order of the model is determined using the auto correlation and partial auto correlation information. M_{11}^{k} and M_{12}^{k} are computed by solving a set of least-square error problem as well. Then, for the i^{th}asset return series, the i^{th} row of M_{11}^{k} and M_{12}^{k} are chosen to minimize

$$\min_{M_{11}M_{12}} || \bar{r}_{i}{}_{k-t}^{k} - (M_{11})_{i}\bar{r}_{i}{}_{k-t-1}^{k-1} - X_{k-t-1}^{k-1}(M_{12})_{i}^{T} ||_{2} \qquad (4.3)$$

Where $i = 1, \ldots\ldots, ||v||_2 \stackrel{\text{def}}{=} \sqrt{v^T v} = \sqrt{\sum_i |v_i|^2}$ is the Euclidean length (vector two norm) for the vector $v = [v1, \ldots]T$, t is the data history length used for estimating M^k, $r_{i}{}_{m}^{n}$ is the return series for asset i over the window $[m, n], X$ is a matrix of macroeconomic time series (one time series per column), and $(M_{12})_{i}$ is the i^{th} row of the sub matrixM_{12}^{k}. Since M^k is upper-triangular, we can compute the macroeconomic factor block M_{22}^{k} independently via the following least squares minimization

$$\min_{M_{22}} || \bar{x}_{j}{}_{k-t}^{k} - \bar{x}_{j}{}_{k-t-1}^{k-1}(M_{22})_{jj} ||_{2} \qquad (4.4)$$

where $(M_{22})_{jj}$ is the jth element of the diagonal matrix M_{22} matrix and $\bar{x}_{j}{}_{m}^{n}$ is the j^{th} macroeconomic time series over the interval $[m, n]$.

4.2.5 Factor-Based Asset Return Model Assumptions

The following fundamental assumptions/simplifications are made for factor-based asset return model:

- M_{11}^k and M_{22}^k are diagonal.
- M_{11}^k and M_{12}^k are determined from macroeconomic time series via the least squares problem defined by equation (4.3).
- M_{22}^k is determined from the macroeconomic time series via the least squares problem defined by equation (4.4).
- There is no coupling from asset return to macroeconomic variables; i.e. M_{21}^k = 0 for all k = 0, 1, ...
- Time series for the following macroeconomic variables are used: Gross Domestic Product (GDP), Consumer Price Index (CPI), Outstanding Consumer Credit, Industrial Production Index, Money Supply (MS), Unemployment Rate, and Ten-Year Treasury.
- One year of data is used for each window (i.e. t = 12 months).
- The assets under consideration consist of sector-based ETFs as well as a risk-free asset.

Future work shall consider relaxing some of the above assumptions.

4.2.5.1 Recursive Lease square

Recursive least square is used when parameters are identified from recurring (in time) linear algebraic equations. The output variable being the sector returns (b) and macro variables forming the input variables (A) matrix,

coefficient (x) are determined in a recursive manner as and when the new data is obtained. The least square solution is given by equation (4.5).

$$\mathrm{Ax} = \mathrm{b} \tag{4.5}$$

where A and B matrix are updated when new data arrives. Updated coefficients x are determined using the equation recursive equation (4.7).

$$\mathrm{A} = \begin{bmatrix} \mathrm{A}_0 \\ \mathrm{a}^{\mathrm{T}} \end{bmatrix}, \mathrm{b} = \begin{bmatrix} \mathrm{b}_0 \\ \mathrm{b} \end{bmatrix} \mathrm{Z} \tag{4.6}$$

$$\mathrm{x} = \mathrm{x}_0 + \mathrm{k}(\mathrm{b} - \mathrm{a}^{\mathrm{T}}\mathrm{x}_0) \tag{4.7}$$

$$\mathrm{k} = \frac{1}{1 + \mathrm{a}^{\mathrm{T}}\mathrm{P}_0\mathrm{a}}\mathrm{P}_0\mathrm{a} \tag{4.8}$$

$$\mathrm{P} = [\mathrm{I} - \mathrm{ka}^{\mathrm{T}}]\mathrm{P}_0 \tag{4.9}$$

Where $P = (A^T A)^{-1}$ and $P_0 = (A_0{}^T A_0)^{-1}$

The recursive formula is initiated by setting P_0 (diagonal matrix) to a to window length of 11 months training period, and x_0 is the initial guess obtained by solving equation {4.5} with A and b data chosen from the training data. This is better compared to the random guess for initial coefficients. A_0 Is not part of the recursive equation making the size of the matrix in the recursive equation to be constant. This is one of the main advantages of this method. The updated coefficients at each step are used to predict the data and entire sequence is repeated to forecast for entire period. Forecasted data is compared with the validation data, to determine the prediction accuracy. The real advantage of using a recursive least square

method is size of the matrix A and B remain constant even when the length of the training window increases at every step.

4.2.5.2 Exponential Weighted Least square

In this method, higher weights are assigned to the latest data when compared to the old data. The weights are assigned based on exponential function .This differs from the moving window method, where the entire length of the window is given equal weight. One could reduce the length of the window such that only the recent past is given equal importance. Reducing the length to such short window will reduce the accuracy of prediction. Steps followed to implement exponential weighted lease square method is discussed below.

- A fixed window length of L is chosen
- The data from the training period of length L is used to fit to the model 1.
- The data is multiplied with the weights generated by the weighting function to assign a decreasing weight starting from recent most value.
- Coefficients/ weights are estimated from the fit.
- These estimated coefficients are used to forecast / predict n steps into future.
- The window is moved, (first data is neglected and new data point is included to fit the model to the training data set), or the window is increased by adding the new data point, and the weights are computed using the exponential weighting function.

- New Coefficients are estimated and are used to forecast the next data.
- In similar manner, entire forecast period is forecasted and validated.
- The window length is changed and steps 1 to step 7 are repeated.

4.2.6 Results

Fig 12, 13, 14 show the sector returns predictions for SPDR. Fig 12 is the mean error % between the predicted monthly sector returns and original monthly sector returns using recursive least square method, where as Fig 13 shows the monthly sector returns accuracy using exponential weighted least square method. However monthly returns are also calculated by computing the average daily returns per month. Fig 14 indicates the worst case error for average daily returns per month predictions.

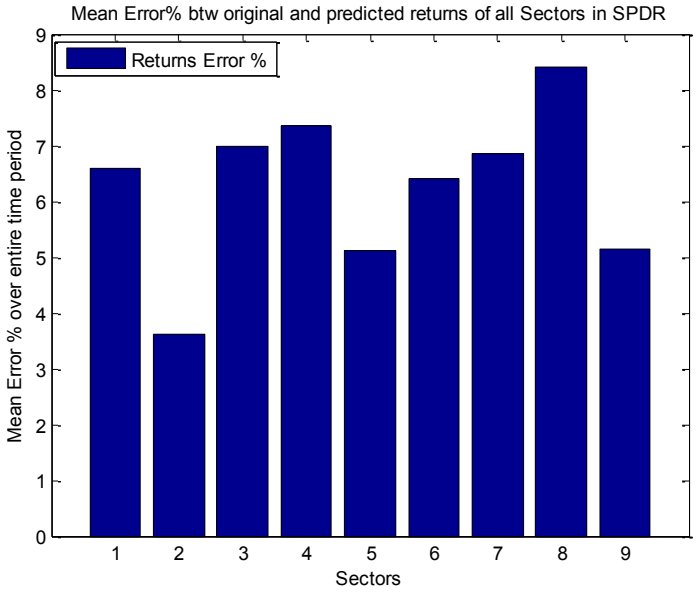

Fig 13**:** Mean error % between original and predicted monthly returns of all sectors in SPDR using regression method

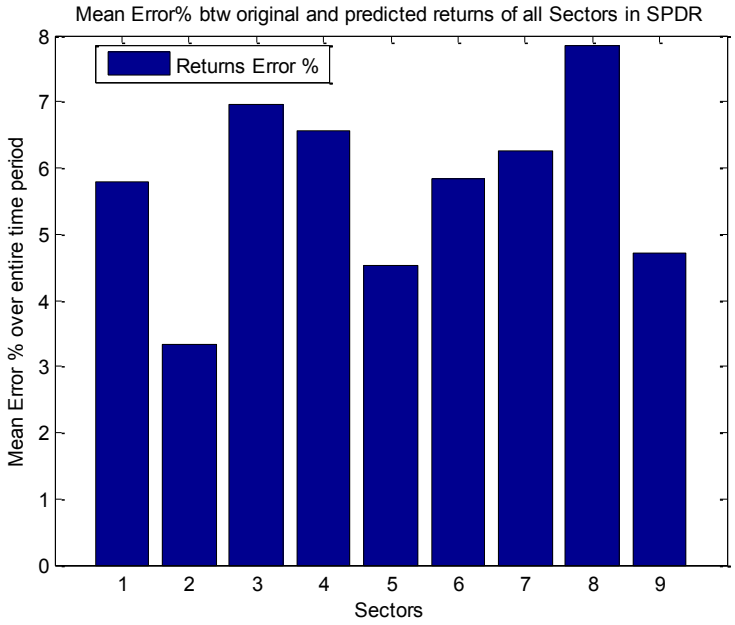

Fig 14: Mean error % between original and predicted monthly returns of all sectors in SPDR using exponential weighted least square method

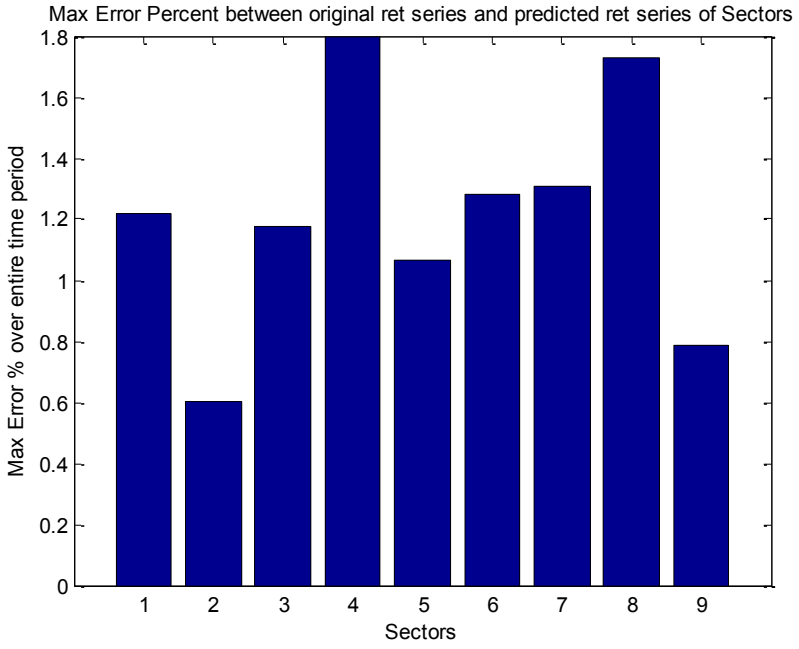

Fig 15 Mean error % between original and predicted average daily returns per month of all sectors in SPDR using regression method

4.3 *Observations:*

- Only important macro series data which has the ability to predict the sector returns are included in the model dynamically.
- There is no significant difference between the different prediction methods.
- Mean prediction error among all the sectors is not greater than 9 %
- Max prediction error among all the sector is higher due to drastic change in the original returns reflecting the change in the economic conditions.
- Average daily returns per month predictions are much better than the monthly return predictions.

5 STOCK/FUND SELECTION

5.1 *Introduction*

Picking the right stock forms the crucial step in the complete investment life cycle. Individual stock returns or combination of stocks return determine the portfolio value at the end of the investment period. Movement of stock prices is based on number of factors. If at all there were handful of them and we knew the effect of each factor on the stock price, then everyone in the market would be billionaires. However that's not the case. Movement of stock price is a complex phenomena which many financial institutions, economist, individual investors and others involved in stock markets try to understand and decode. Over the years, many models have been presented which tried to capture the underlying phenomena of stock price movement. Every model has some limitation leaving the puzzle to be solved. Two widely used techniques/methods to rightly pick the stocks are Fundamental analysis and Technical analysis. Stock valuation models determine the real price of a stock, with the fundamental data and other exclusive company information which are helpful in stock picking decision.

Different fundamental ratios are explained and methods to indentify the key fundamental ratios which help in picking the right stock is presented. Technical analysis and stocks selection for optimization are not addressed in this thesis and will be considered for future work.

5.2 *Fundamental analysis*

Fundamental analysis involves market analysis, industry analysis, company analysis and portfolio management (Frank K. Reilly 2002). Focusing on the company analysis, the company's balance sheet, income statement and cash

flow statement are studied to determine the health of the company and use that information to predict the performance of the stock. Balance sheet, income statement and cash flow statement provide all the required numbers to compute the fundamental ratios. These ratios are formed and analyzed which help in understanding the company future performance.

Fundamental ratios are further classified into 6 different categories such as Liquidity measurement ratios, Profitability indicator ratios, Debt ratios, Operating performance ratios, Cash flow ratios and Investment valuation ratios (Investopedia n.d.). Some of the key ratios which are widely analyzed are Price/Earnings ratio, Price/Book value, Market cap, Price to sales ratio, Dividend growth, Operating cash flow, and Debt ratio. Apart from the mentioned ratios other important ratios such as cash ratio, quick ratio, Return on Assets, Return on Equity, debt/equity ratio etc are computed and used in the analysis.

5.2.1 Fundamental ratios

Key fundamental ratios considered in this thesis are categorized and defined. These ratios can be used to compare different companies and identify the best among them. However caution must be exercised when comparing companies from different industries. Important ratios for each industry has been presented in section {5.4.3}. Key fundamental ratios are defined below.

5.2.2 Liquidity measurement ratios:

These ratios indicate the ability of company to pay its short term debt obligations. They are calculated using the most liquid assets in companies possession.

5.2.2.1 - Current Ratio

It's the ability of the firm to pay off its current liabilities using its current assets. It's the ratio of current assets to current liabilities. Higher the ratio better the firm.

5.2.2.2 - Quick Ratio

It's a conservatory but a similar measure than current ratio. Inventories are excluded from the current assets. Hence it's the ratio of current assets (excluding inventories) to current liabilities.

5.2.2.3 - Cash Ratio

Further refined measure of quick ratio which includes only cash, cash equivalents and accounts receivable in current assets. Higher the ratio, better the company.

5.2.2.4 - Cash Conversion Cycle

It's the time taken in days by the firm to liquidate its current assets including inventories, collect all the receivables and pay its current liabilities. It indirectly determines the strength of the firm in terms of its working capital. The shorter the duration to sell its inventories, collect its receivables shows the liquid nature of the firm.

CCC = DIO + DSO – DPO

CCC – cash conversion cycle

DIO – Days inventory outstanding

DSO- Days sales outstanding

DPO- Days payable outstanding

5.2.3 Profitability indicator ratios:

These ratios indicate the performance and profitability of a company and its ability to manage its resources efficiently in generating profits and value to its shareholders.

5.2.3.1 - Profit Margin Analysis

Profit margins are computed from income statement. Gross profit, operating profit, pre-tax profit and net profit are four different profit numbers reported in an income statement. These numbers are used to compute the profit margins. Ratio of profits to net sales (revenue) indicates the quality of companies' investment and their growth. Trends in the profit margins are analyzed to determine the profitability and performance of the company.

5.2.3.2 - Return on Assets (ROA)

Widely used measure for stock valuation is return on assets which indicates the ability of the firm to generate profits from the total assets available. It's the ratio of net income to average total assets. Higher number is preferred.

5.2.3.3 - Return on Equity

It's a good indicator to show how much an investor has earned from the company performance over a specific period, typically every quarter. It's the ratio of net income to average share holders' equity. Higher number indicates increase in shareholders values and the investment in that firm is considered worthy.

5.2.3.4 - Return on Capital Employed

ROCE is the ratio of net income to the capital employed. it's an aggressive measure, which gauges the firm's ability to generate profits from its capital.

5.2.4 Debt ratios:

Debt ratio indicates the level of risk companies and its shareholders face. It gives the back ground information about the firm. Debt is defined as short term borrowings, cash payable, current portion of long term borrowings and long term borrowings.

5.2.4.1 - Debt Ratio

Ratio of total liabilities to total assets. Indicates the ability to the firm to operate without having to borrow money. Smaller the ratio lesser the company is dependent on debt.

5.2.4.2 - Debt-Equity Ratio

It provides an idea on the firms leverage position. It's the ratio of total liabilities to shareholders equity. Total liabilities include, debts from private investors, supplies, lenders. Shareholder equity is amount invested by the shareholders. Lower the numbers better the health of the firm.

5.2.5 Operating performance ratios:

Operating ratios gives an insight into the ability of the firm to converts its assets to revenues and sales to cash. Better the ratios better it is for the shareholders.

5.2.5.1 - Fixed-Asset Turnover

Fixed asset turnover is a rough estimate of the productivity of the firm. It's the ratio of revenue to fixed asset value which include, plant, property and equipment (PP&E).

Higher the number better the company. It's a sector and industry dependent number and is not appropriate to use in comparison between two companies from different industry.

5.2.5.2 - Sales/Revenue Per Employee

Its also an productivity measure, but from its work force. Indicates the management ability to balance the required work force and increase its revenue using them. It's the ratio of revenue to total number of employees.

5.2.6 Cash flow ratios

Cash flow ratios indicate the safety net a firm has from the cash being generated.

It's the cash generated from the sale of the products that can be used for operations and to pay short term debts. Different cash flow ratios are explained.

5.2.6.1 - Operating Cash Flow/Sales Ratio

Ability to turn sales to cash is measured (in percentage) by operating cash flow to sales (revenue) ratio.

5.2.6.2 - Free Cash Flow/Operating Cash Ratio

It's the ratio of free cash flow (operating cash flow – capital expenditure) to operating cash flow. Higher the percentage of free cash flow greater is the financial strength of the firm.

5.2.6.3 - Dividend Payout Ratio

It is the ratio of dividends per common share to earnings per share expressed in percentage. It indicates how well the earnings support the dividend payment.

5.2.7 Investment valuation ratios

Important and widely used ratios for stock valuation are the investment valuation ratios. These ratios are computed, analyzed and shared by financial advisors, companies and business media. It indicates the attractiveness of the potential investment or existing investment. All the above ratios computed using the income statement, balance sheet and cash flow statement are complex to understand and compute the required fundamental ratio. Investment valuation ratios simply the process and incorporate the important information for valuation and investment. These investment ratios are used in factor models.

5.2.7.1 - Price/Book Value Ratio

Price to book value is widely used valuation measure indicates the amount an investor is paying for the net assets of the company. It's the ratio of stock price per share, to shareholders equity per share. Lower ratio indicates that the investor is paying less for what the company is worth.

5.2.7.2 - Price/Earnings Ratio

It's the single most widely used ratio which indicates the amount investors are ready to pay per dollar of earning. it is the ratio of market value per share to earnings per share (EPS). EPS is calculated from the balance sheet, which is ratio of net revenue to total shares outstanding. Higher P/E suggest that investors are expecting higher returns and hence ready to pay more compared to low P/E companies.

5.2.7.3 - Price/Cash Flow Ratio (PCF)

PCF is ratio of stock price per share to operating cash flow per share.

5.2.7.4 - Price/Earnings to Growth Ratio (PEG)

PEG is closely related measure to P/E ratio. It's the ratio of P/E ratio to Earning per share growth. When the ratio is less than 1, future earnings are higher than the current valuation of stock. However PEG=1 indicates that the current valuations are right. EPS growth is forecasted and used for current valuation.

5.2.7.5 - Price/Sales Ratio

It's the ratio of stock price per share to net sales per share. It is similar to P/E ratio, where price of stock against net sales is evaluated instead of earnings. It's the measure of money an investor is willing to pay for every dollar of company's sales.

5.2.7.6 - Dividend Yield

This ratio is important for income investors, where the income obtained from dividend per share is compared with stock price per share for valuations. It's the ratio of dividends per share to stock price per share.

5.3 *Restriction of asset universe for further analysis/optimization:*

Thousands of stocks are traded on the stock exchanges such as NASDAQ and New York stock exchange. Once the right sector is identified based on sector analysis {4.2}, the select set of stocks to invest are identified. This leads to the problem of reducing the stock universe of size greater than 1000 to 5 or 10 which would be part of the portfolio.

All stocks are classified into Mega cap, large cap, mid cap and small cap. These classifications are based on their market capitalization which is defined as, total dollar market value of a company's outstanding shares (Investopedia n.d.). It calculated by taking the product of current day share price with the total number of outstanding shares. (The number of outstanding shares is obtained from the annual/quarterly reports filed by the company).

Different methods are employed to reduce the stock universe. Stock valuation is a widely used methodology which determines the actual stock price based on the fundamentals and analyst predictions. Difference between the predicted price and the market price is used in the decision process. Investors typically tend to pick the top 5 performing stocks and include them in their portfolio. Important parameter that needs to be considered in the stock filtering process is discussed in section {5.5}

5.3.1 Stock price factor determination

In this section the fundamental factors which aid in determining the actual stock price is discussed. Stocks in different sectors in different time periods are considered.

5.3.1.1 Method: Correlation Analysis

Commonly used method to determine the relation between the fundamental ratios and the stock price/stock returns is correlation analysis. Correlation explains the strength and the direction of the linear relation between two different variables. The correlation coefficient is between -1 and 1. When the linear relation is higher between the different variables then the value is towards 1 and when its weaker then its towards -1. The generic correlation coefficient equation is given as

$$\rho_{x,y} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sigma_x \sigma_y} \qquad (5.1)$$

Where

E - Expected Operator

μ - Mean of the series

σ - Standard deviation of the series

The correlation coefficient value is interpreted as the ability of a particular factor to predict the stock price or stock returns. The returns of a particular stock and not the stock price itself is considered to compute the correlation coefficient between the fundamental ratios such as PE, PES, PEG etc with the stock returns.

Stock returns at k+1 time period are correlated with the fundamentals obtained from balance sheet, income statement and the cash flow statement at time period k, value of which is a good indicator of the predictive ability of the fundamental ratios.

5.3.1.2 P-value

Correlation coefficient determines the relation between two different variables, but fails to determine, the significance of the relation. The objective is to identify the factor which has greater influence on the stock returns and the ability to predict the returns. P-value with range 0 and 1 is used to determine the significance of a relation between different variables. It determines the probability of hypothesis being true. If the hypothesis is that, there is no significant relation between the fundamental ratios and the stock returns. Value <0.05 (0.05 number is chosen based on 1 standard deviation.) indicates that the probability of the hypothesis being true is less than 5% which indirectly signifies, 95% of the time, the hypothesis is incorrect and hence can be rejected which confirms the significance of the relation between the fundamental ratio and the stock returns.

5.3.2 Data and Implementation

Fundamental ratios vary from month to month, Quarter to Quarter and year to year. Hence correlation coefficient computed at one point does not represent the relation between the returns and the fundamental ratios all the time. It also changes with sector and company. To obtain a realistic relation between the ratios and returns to pick the right stock, the entire time period from 1987 to 2006 is divided into 4 different regimes.

- 1993-1999 and 2003-2006 - Economic boom period
- 1987-1992 and 2000-2002 - Recession

All the companies are classified under 8 different sectors such as Basic Materials, Consumer goods, utilities, technology, financial, services, industrial productions, conglomerates', other indices. Based on the classification the correlation coefficients between the fundamental ratios and month ahead stock returns are computed. SP500 market index which represents the entire market is considered as a stock.

5.3.3 Results

Table 6 - Table 14 show the correlation coefficient between the various fundamental ratios of all the companies in a particular sector and the SP500 month ahead index returns. Only values with P-value < 0.05 are considered. These fundamental ratios can be used in the stock price prediction models.

5.3.3.1 Basic Materials

This sector consists of companies involved in the raw material business. They are expected to perform well during good and the bad times of the economy based on the industry segment which they belong to. They heavily depend upon the availability of the raw material/resources for the survival. E.g. Oil companies have constant demand; however they are always dependent upon the crude oil availability. Hence they should always be having enough amounts of cash to bid for new reserves and also for exploration in the new areas. Hence **Liquidity ratios** play an important role in determining how well the company is performing. **Profitability ratios** also give an insight on performance of the company.

Table 6: Correlation Coefficient between SP500 index returns and fundamental ratios in basic materials sector

Liquidity Measurement	*87-2006*	*87-92*	*93-99*	*2000-2002*	*2003-2006*
Days of Payable Outstanding	0.05		0.079	0.11	0.074
Cash Conversion Cycle	0.051		0.083	0.102	0.075
Cash ratio	0.08		0.072		0.128
Quick Ratio	0.076				0.131
Current Ratio	0.055				0.138
ROA (Profitability ratio)				0.183	-0.116
Price/Sales (Evaluation ratio	0.082		0.183	0.119	0.121

5.3.3.2 Conglomerates

These are the big companies which have its branches in many industry segments. They have portfolio of products which cater different needs. From the correlation analysis we can see that even here liquidity ratios play a major role in giving us an insight on the performance of the companies. Apart from their **liquidity** ratios their return on asset operating profit margin and their **debt** ratios seems to tell us how they perform during the bearish period.

Table 7: Correlation Coefficient between SP500 index returns and fundamental ratios in conglomerates sector

Liquidity Measurement	*87-2006*	*87-92*	*93-99*	*2000-2002*	*2003-2006*
Days of Inventory Outstanding		0.245	-0.212		
Days of Sales Outstanding		0.206	-0.224		-0.532
Days of Payable Outstanding			0.148	-0.53	
Cash Conversion Cycle				-0.528	-0.56

Cash ratio					0.43
Quick Ratio					0.427
Current Ratio			-0.213		
Profitability Ratios					
Return on Capital Employed		0.198		0.396	
ROE		0.236			
ROA				0.489	
Gross Profit Margin					
Operating Profit Margin				-0.478	
Debt ratios					
Debt Ratio			0.173	-0.612	
Debt Equity Ratio			0.155	-0.609	
Capitalization ratio				-0.61	
Valuation ratios					
Market Cap			0.147	-0.548	-0.393
Enterprise value				-0.709	

5.3.3.3 Consumer Goods

This sector consists of all the industries which fulfill the common mans day to day requirements, like food, beverages, clothing, office supplies and etc. These companies have evergreen demand. Their performance is measured by how much volume they are selling to make enough profits. Hence all the ratios which are based on the **revenue** and **sales** give us a good insight on their performance.

Table 8: Correlation Coefficient between SP500 index returns and fundamental ratios in consumer goods sector

Investment Valuation Ratios	87-2006	87-92	93-99	2000-2002	2003-2006
Earnings Per share	0.082			0.092	0.183
Price/Book Value		0.09			
Price/cash Flow ratio	0.053	0.098			

Price/Earnings					
Price/Sales			0.083		-0.1
Market Cap			0.093	-0.106	
Enterprise value			0.123		
Cash Flow Indicator Ratios					
Capex+Cash Dividend Coverage	0.11	0.165	0.093	0.116	

5.3.3.4 Financial

This sector consists of all the companies which are dealing with money which includes Banks, insurance, equity managements and etc. Their performance largely depends on the amount of **cash/ cash flow** they have for their operations. Effective management of their available cash indicates a good performance of the company. Hence cash flow ratios are important when we are trying to analyze the financial companies. However the correlation analysis doesn't confirm our hypothesis. The ratios which are important for the stock returns are the **valuation ratios.**

Table 9: Correlation Coefficient between SP500 index returns and fundamental ratios in financial sector

Investment Valuation Ratios	87-2006	87-92	93-99	2000-2002	2003-2006
Earnings Per share		0.186	0.168		
Price/Book Value	0.474	0.143			0.188
Price/cash Flow ratio					
Price/Earnings	0.12	0.238			0.108
Price/Sales		0.191	0.078		
Market Cap		0.205			
Enterprise value		0.225			

5.3.3.5 Health care

This sector consists of all the companies which are related to the health care. Drug manufacturing companies, hospitals, biotechnology and etc. **Profit** margins play a key role in their performance measurement. Typically drug companies depend on the patents, licenses for drug and they spend lot of their income on Research. Hence management of **debt** becomes an important criterion. However they are greatly dependent on the kind of policies the government puts in place. Hence we need to look into some other ratios based on our knowledge on the government policies.

Table 10: Correlation Coefficient between SP500 index returns and fundamental ratios in health care sector

Profitability Indicator Ratios	87-2006	87-92	93-99	2000-2002	2003-2006
Return on Capital Employed		0.363		0.203	-0.124
ROE		0.358		0.161	-0.138
ROA				0.172	-0.156
Gross Profit Margin				-0.128	
Operating Profit Margin				0.114	
Net Profit Margin					-0.273

5.3.3.6 Services

This sector consist of wide range of companies which provide day today services such as TV, auto, sports, media, transport and etc. All these primarily depend on the number of customers they have each and every day/week or month. Hence their profitability margins give us a good insight on how they are performing. This is because their investment is not

continuous, it all depends on how well they provide their service and how many people they are able to attract or retain.

Table 11: Correlation Coefficient between SP500 index returns and fundamental ratios in Services sector

Profitability Indicator Ratios	87-2006	87-92	93-99	2000-2002	2003-2006
Return on Capital Employed			0.141		
ROE			0.27		
ROA	-0.063		0.175	0.116	-0.191
Gross Profit Margin					
Operating Profit Margin					
Net Profit Margin	-0.036		0.095	0.086	-0.122

5.3.3.7 Technology

This sector includes all the companies which are one way or the other related to the technology. Most of them are product manufacturing companies. They require good amount of money for research and development to keep coming up with new products. Here again they performance measure is based on how much volume they are able to sell and how much revenue they are able to produce. Hence all the ratios based on the revenue would be a good indictor of their performance. During bad times they suffer due to the dip in the sales. During those times their liquidity ratios play a key role, because the ability of the companies to be floating and sustaining themselves is of prime importance.

Table 12: Correlation Coefficient between SP500 index returns and fundamental ratios in Technology sector

Investment Valuation Ratios	87-2006	87-92	93-99	2000-2002	2003-2006
Earnings Per share	0.061				
Price/Book Value	0.129	0.191	0.169		
Price/cash Flow ratio	0.116		0.41		
Price/Earnings	0.103	0.182	0.107		0.168
Price/Sales	0.236	0.134	0.355		0.105
Liquidity Measurement					
Cash ratio	0.061		0.191	-0.103	
Quick Ratio	0.047		0.185	-0.112	
Current Ratio			0.154	-0.106	
Operating Performance Ratios					
Fixed Asset Turnover	0.07	0.151	0.074	0.186	

5.3.3.8 Industrial Goods

This sector primarily consists of industries and companies which cater the basic needs of the consumer. The performance of this sector is again dependent on other sectors such as consumer goods, services, utilities. Few industries perform better during bad times and worse during good times, based on the type of industry segment they are in. Mostly their **profitability** ratio and the ability to maintain/payoff their **debts** give us an insight on how well they perform.

Table 13: Correlation Coefficient between SP500 index returns and fundamental ratios in Industrial Goods sector

Debt Ratios	87-2006	87-92	93-99	2000-2002	2003-2006
Debt Ratio			0.107	0.115	0.246
Debt Equity Ratio				-0.147	0.097

Capitalization ratio			0.142	0.188	
Cash to debt Ratio				0.107	
Profitability Indicator Ratios	87-2006	87-92	93-99	2000-2002	2003-2006
Return on Capital Employed			0.909		
ROE			-0.156	0.106	-0.099
ROA		-0.099	0.245	0.273	-0.208
Gross Profit Margin					-0.152
Operating Profit Margin		0.181		0.097	-0.103
Net Profit Margin				0.232	-0.176

5.3.3.9 Utilities

This sector is again similar to the consumer good sector where companies address our day today requirements for survival. This includes electricity, gas, water and waste management. They are greatly dependent on how many customers they have, and how efficiently they are operating. Since there is a constant and an increasing demand day by day, their performance measure depends on many ratios such as the **profitability** margins, **cash** flow ratios and amount of liquidity they have and etc. Hence almost all the ratios except of debt give us a insight on the performance.

Table 14: Correlation Coefficient between SP500 index returns and fundamental ratios in Utilities sector

Investment Valuation Ratios	87-2006	87-92	93-99	2000-2002	2003-2006
Earnings Per share	0.163	0.314	0.236	0.434	0.13
Price/Book Value		0.174		0.448	
Price/cash Flow ratio		-0.164			
Price/Earnings					
Price/Sales	0.058	0.113		0.232	
Market Cap				0.162	

Cash Flow Indicator Ratios					
OCF/Sales		0.112			
FCF/OCF		0.237			
capital Expenditure coverage	0.098	0.194			0.16
Short Term Debt coverage					
Dividend Coverage	0.062	0.178			0.183
Capex+Cash Dividend Coverage	0.148	0.242	0.1		0.193
Profitability Indicator Ratios					
Return on Capital Employed	0.068		0.386	0.391	
ROE			0.214		
ROA	0.227	0.263	0.443	0.376	0.201
Gross Profit Margin				0.149	
Operating Profit Margin					
Net Profit Margin	0.238	0.252	0.41	0.311	0.252

5.4 *Obtaining bounds/thresholds on fundamental values to predict performance*

The fundamental ratios which are significant in predicting the future stock returns in a particular sector were identified in section 5.3 . It is extremely difficult to analyze each and every stock in each sector. It's a humungous task to read the income statement, balance sheet and cash flow statement of all the stocks and separate the good and the bad ones based on the fundamental ratios. Hence an automatic screener called filters which would filter out the stocks based on the user defined constraints is used. Those stocks which satisfy the user criteria alone will make it to the final list. In order to indentify the bounds/threshold on fundamental values below steps are followed.

- Companies are categorized as Mega Cap, Large Cap, mid cap and small cap. Companies whose market cap is greater than 150 billion are tagged as Mega cap, between 50 billion and 150 billion as large cap, greater than 5 billion and less than 50 billion as mid cap, and all the companies below 5 billion are small cap. This categorization differs from analyst to analyst.
- The entire time period under study is divided into boom and bust periods based on historical performance.
- The benchmark (SP500) returns are computed annually for the entire time period.
- The stock universe is reduced based on benchmark returns, and/ or user defined constraints.

5.4.1 Data

The stock price data is obtained from yahoo finance, by downloading the daily stock close, open, high, low price along with the volume data. Apart from the ratios identified in Table 6-Table 14 which are significant in predicting the stock returns , few other common ratios such as Sales Per employee, Return on Asset , Return on Equity and Net profit margin are also included. Fundamental data is obtained from Compustat database. The analysis is performed on different sectors, with the similar categorization as mentioned below.

- Sectors :Technology, services, Utilities and etc
- Time Periods: 1992- 1999, 2000-2002, 2003-2006
- Category of stocks:
    - Mega Cap – Market cap greater than 150 Billion

- Large Cap – Market cap between 50 billion and 150 billion
- Mid Cap - Market cap between 5 billion and 50 Billion

- Benchmark returns: Average S&P 500 index returns for the different time periods

5.4.2 Methodology

5.4.2.1 Method 1: Filters based on Benchmark returns

- Filter 1: Time window (e.g. : 1993-1999)
- Filter 2: Market cap greater than Y (or between Y1 & Y2) – Y1 and Y2
- Filter 3: Average returns greater than x% (x- determined based on time and benchmark returns)

5.4.2.2 Method 2: Filters based on User defined constraints

Constrains or the RULES of THUMB values based on users knowledge or analyst recommendations for the fundamental ratios.

- Filter 1 :Same as Method 1 Filter 1
- Filter 2 : PE =< 30, ROA>5%, ROE>15%
    - ROA- Return on Assets should be in greater than 5%.
    - ROE- Return on equity should be greater than 15-20%
    - Price to Earnings ratios should be less than 30
- Filter 3 : Market cap greater than Y (or between Y1 and Y2)

Values Computed:

- Average: Mean values of the fundamental ratios are computed for those companies which are shortlisted on the above mentioned criteria.
- Min and Max: Minimum and Maximum value among the values

- Exceptions: Those values which are extremely away from the average value and change the average greatly when included in computing the average value which leads to the misinterpretation of the ideal value. These values include extreme low and high values.

5.4.3 Results

5.4.3.1 Method 1

Filters are applied on the historical fundamental data based on benchmark returns Filter 1:

- Market cap greater than 150 Billion (Mega Cap)
- Market cap > 50 Billion & < 150 Billion (Large Cap)
- Market cap > 5 Billion & < 50 Billion (Mid Cap)

Filter 2:

Mega Cap:

- 1993-1999 - Average returns greater than 15% (S&P500 = 19%)
- 2000-2002 - Average returns greater than -20% (S&P500 = -15%)
- 2003-2006 - Average returns greater than 13% (S&P500 = 15%)

Large Cap:

- 1993-1999 - Average returns greater than 20% (S&P500 = 19%)
- 2000-2002 - Average returns greater than -15% (S&P500 = -15%)
- 2003-2006 - Average returns greater than 15% (S&P500 = 15%)

Mid Cap:

- 1993-1999 - Average returns greater than 25% (S&P500 = 19%)
- 2000-2002 - Average returns greater than -10% (S&P500 = -15%)
- 2003-2006 - Average returns greater than 20% (S&P500 = 15%)

The expected returns are greater and lesser the average benchmark returns based on the category of stock. Mega cap companies are large and provide stability to the portfolio. Growth in these companies are lesser when compared to the mid cap companies. Hence the expected returns are less than the average returns of the benchmark. Large Cap companies are big and they are considered to be the market leaders and indicators. Hence the expected returns are greater or equal to the benchmark returns. Mid Cap companies are small and have an excellent growth story. There is considerable amount of risk associated with them, but their returns are higher than large cap and mega cap. Hence they are expected to beat the benchmark with a margin >= 5%.

Table 15: Fundamental Values of the Mega Cap Companies- Method 1

	1993-1999			2000-2002			2003-2006		
Fundamental	(6 entries)			(4 entries)			(5 entries)		
Ratios	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
PE	43	24	85	31	20	48	32	20	47.3
EPS	3.5	0.4	6.9	2	0.2	4.4	1.59	0.5	4.47
PBV	7.3	5	9	6.8	5.3	8.8	5.13	1.2	7.03
PS	7.8	2	15.5	4.4	2.4	7.9	5.3	1.7	8.98
Current Ratio	1.5	0.5	2.3	2.05	0.76	3.5	0.89	0.1	2.36
Quick Ratio	0.8	0.03	2.21	1.4	-0.03	2.75	1.35	0.1	2.36
Fixed asset turnover	4.7	0.6	10.8	2.4	1	5.2	4.39	0.6	8.28
Sales/Emp	292	204	410	340	233	538	401	208	581
ROA	10	1.32	19	8	2.91	12	8.89	2.7	12.8
ROE	21.97	3.9	34	24	15.5	32.1	16.9	6.3	27.1

Net profit margin	12%	4%	23%	15%	5%	29%	15%	9%	20%

Complete list of companies are given in Appendix A1

Companies shortlisted:

- 1993-1999:
    - Total Number of Companies : 9
    - Companies satisfying the constraints: 6 (Oracle, IBM, Intel. A total of 6 entries (Entry = one company in one particular year)
- 2000-2002:
    - Total Number of Companies : 7
    - Companies satisfying the constraints : 4 (Microsoft, IBM, Intel, AT&T)
- 2003-2006:
    - Total Number of Companies : 5
    - Companies satisfying the constraints : 5 (CISCO, INTL, IBM, MS, AT&T)

Table 16: Fundamental Values of the Large Cap Companies- Method 1

	1993-1999 (25 entries)			2000-2002 (5 entries)			2003-2006 (17entries)		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
PE	53	11	116	47.5	32.5	56.7	26.7	12.3	48.1
EPS	3.25	0.2	10.9	0.52	0.12	1.04	2.02	0.37	6.37
PBV	10.5	1.7	36	8.9	3.9	15	5.81	1.81	16.3

PS	6.27	0.6	17.1	2.09	1.91	2.24	2.48	0.93	6.73
Current Ratio	1.79	0.1	5.7	1.19	0.6	2.14	1.54	0.59	3.91
Quick Ratio	1.17	0.07	4.4	0.6	0.02	1.4	0.79	0.13	2.27
Fixed asset turnover	5.5	0.6	14.2	2.71	0.5	6.7	8.89	1.07	29.0
Sales/Emp	352	178	588	519	225	916	461	244	965
ROA	11.1	1.3	21.7	7.49	3.08	13.7	9.07	1.08	21.0
ROE	20	2.8	34.2	10.5	1.2	26.5	21.0	4.21	46.9
Net profit margin	14%	2%	38%	5%	1%	12%	10%	2%	26%

Complete list of companies are given in Appendix A1

Companies shortlisted:

- 1993-1999:
    - Companies satisfying the above constraints : 21 (CISCO, HP, MS, IBM and others) among a total of 25 companies
- 2000-2002:
    - Companies satisfying the above constraints: 4 (DELL, Verizon, Alcatel, Taiwan Semiconductors) among a total of 20 companies.
- 2003-2006:

- Companies satisfying the above constraints : 12 (Apple, DELL, MOTO, Oracle and etc) among a total of 20 companies

Table 17: Fundamental Values of the Mid Cap Companies- Method 1

	1993-1999 (110 entries)			2000-2002 (42 entries)			2003-2006 (87entries)		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
PE	52	10.56	342	96	-307	2451	43	8.8	131.2
EPS	2.19	0.03	7.19	0.6	-3.3	3.4	1.15	0.04	3.98
PBV	9.5	1.2	64.3	5.4	1.48	14.05	4.65	1.45	13.29
PS	6.2	0.43	54.12	5.5	1.1	19.88	4.3	0.44	13.93
Current Ratio	2.25	0.5	6.24	2.54	0.5	10.26	2.27	0.47	7.36
Quick Ratio	1.07	-0.3	6.03	1.26	-0.3	4.03	1.49	0.03	9.37
Fixed asset turnover	5.38	0.14	22.33	5.7	0.3	15.3	4.96	0.2	17.89
Sales/Emp	288	95	1020	301	79	829	353	108	843

ROA	10.1	1.16	29.9	4	-21	20	7.81	-72	66
ROE	16.36	-8	47	8.6	-40	33.9	4.15	-47	23
Net profit margin	11%	-8%	38%	6%	-37%	21%	9%	-8%	27%

Complete list of companies are given in Appendix A1

Companies shortlisted:

- 1993-1999:
    - Companies satisfying the above constraints : 74 (Microsoft APPL, NORTELL, VERIZON, 3COM, ALCATEL, VERIZON, DELL, EMC, KYOCERA and others) among a total of 86 companies.
- 2000-2002:
    - Companies satisfying the above constraints : 46 (Lexmark, First data corp, appl, amd, qlogic, sprint, nvidia, LSI and others) among 92 companies
- 2003-2006:
    - Companies satisfying the above constraints : 67 (AMD, APPL, Nortel, yahoo, Lexmark, Juniper and others) among 78 companies

5.4.3.2 Method 2

Filter 1:

- Market cap greater than 150 Billion (Mega Cap)
- Market cap > 50 Billion & < 150 Billion (Large Cap)

- Market cap > 5 Billion & < 50 Billion (Mid Cap)

Filter 2: PE =< 30, ROA>5%, ROE>15%

Filter 3:

Mega Cap:

- 1993-1999 - Average returns greater than 15% (S&P500 = 19%)
- 2000-2002 - Average returns greater than -20% (S&P500 = -15%)
- 2003-2006 - Average returns greater than 13% (S&P500 = 15%)

Large Cap:

- 1993-1999 - Average returns greater than 20% (S&P500 = 19%)
- 2000-2002 - Average returns greater than -15% (S&P500 = -15%)
- 2003-2006 - Average returns greater than 15% (S&P500 = 15%)

Mid Cap:

- 1993-1999 - Average returns greater than 25% (S&P500 = 19%)
- 2000-2002 - Average returns greater than -10% (S&P500 = -15%)
- 2003-2006 - Average returns greater than 20% (S&P500 = 15%)

Table 18: Fundamental Values of the Mega Cap Companies – Method 2

	1993-99 (2 entries)	2000-2002				2003-2006 (9 Entries)		
		4 Entries (without filter 4)			2 entries (with filter4)			
	Mean	Mean	Min	Max	Mean	Mean	Min	Max
PE	25	22	19	26	23	24.3	19.2	29.6
EPS	4.5	2.5	1.5	4.4	3.4	2.23	0.9	5.12

PBV	16	6.8	5.3	8.8	7	6.13	4.8	7.71
PS	8.7	5.3	2.4	9.9	2.7	5.37	1.68	9.15
Current	1.5	1.8	0.76	3.1	0.9	1.8	0.44	4.21
QR	1.1	1.02	-.03	2.3	0.15	1.49	0.3	3.76
Fixed asset turnover	7.5	2.4	1	5.2	3.2	11.12	6.06	16.95
Sales/Emp	263	370	233	588	251	502	279	652
ROA	27	14	8	21	8.5	12.62	7.26	18.1
ROE	65	27	22	32	29	25.36	16.8	31.4
Net Profit margin	35%	24%	9%	41%	12.24%	21%	9%	31%

Complete list of companies are given in Appendix A1

Companies shortlisted:

- 1993-1999:

  Companies satisfying the constraints: 2 (Oracle, IBM) out of 9 companies.

  - INTL is filtered out when compared with Method 1 and Only 2 entries remained each for only 1 year in the period of 7 years.

- Above calculated values do not truly represent the good fundamentals of all the companies above 150 billion

- 2000-2002:

  Companies satisfying the constraints: 3 (Microsoft, IBM, Intel) out of 3 co's. However When Filter 3 is applied, data is insignificant to compute the values.

- 2003-2006:

  Companies satisfying the constraints: 3 (CISCO, MSFT, IBM) out of 5 companies

Table 19: Fundamental Values of the Large Cap Companies –Method 2

	1993-1999 (6 Entries)				2003-2006(5 Entries)		
	Mean	Min	Max		Mean	Min	Max
PE	17	12	20		18.5	12	25
EPS	6.8	2.8	10.9		2.2	0.51	5.2
PBV	4.2	2.3	6.7		4.7	3	8.6
PS	2.9	0.69	5.8		2.9	1.2	6.7
Current	1.5	0.9	2.8		2.12	0.62	3.9
QR	0.6	0.2	2		1.06	0.26	1.8
Fixed asset turnover	3.2	0.83	6.79		8.8	1.5	16
Sales/Emp	304	178	434		421	247	587

ROA	11.4	5.2	21.73		12.5	5.72	21
ROE	23.93	18.63	30.51		26	16	33
Net profit margin	17%	6%	38%		15%	7%	26%

Complete list of companies are given in Appendix A5

Companies shortlisted:

- 1993-1999:

  Companies satisfying the constraints are 5 out of 25 which includes IBM, HP and INTL.

  - Big players such as MSFT, AT&T, MOTOROLA, Sun etc were filtered out
  - Range of values has decreased
  - Variation from the average value is much smaller compared to method 1

- 2000-2002:

  Companies satisfying the above constraints : 0 out of 20

  - During recession- method 1 is better than method 2.

- 2003-2006:

  Companies satisfying the constraints: 6 (HP, TI, MOTOROLA, ORACLE and others) out of 20. Over a period of 4 years, not a single company meets the constraints for more than 2 years and total number is reduced when compared to method 1.

Table 20: Fundamental Values of the Mid Cap Companies – Method 2

	1993-1999 (28 entries)			2000-2002 (6 entries)			2003-2006 (20 entries)		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
PE	20	10.5	29.8	20	10	28	18.8	8.8	25.02
EPS	3.4	1.5	7	2.8	2.09	3.4	2.52	0.79	6.5
PBV	5	1.9	11	5.4	3.4	7.8	5.4	2.5	8.2
PS	2.3	0.6	4.2	2.5	1.7	3.6	2.7	1.2	5.6
Current	1.7	0.5	4.2	1.6	1.1	2.4	1.7	0.6	3.8
QR	0.7	-0.1	2.3	0.5	-0.3	1.06	0.8	0.03	3.5
Fixed asset turnover	5.7	0.6	19	4.7	0.9	9.1	3.5	1	9
Sales/Emp	324	115	775	311	211	364	382	98	1341
ROA	12.5	6.6	27.8	9.6	5.3	13.06	13	7	20
ROE	27	15	73	26	19	33.9	30	16	66
Net profit margin	13%	5%	29%	13%	7%	17%	15%	6%	25%

Complete list of companies are given in Appendix A6

Companies shortlisted:

- 1993-1999:

  Companies satisfying the constraints are 23 out of 86. Companies include Microsoft,APPL,NORTELL,VERIZON,3COM,ALCATEL,VERIZON,DELL,EMC,KYOCERA

  and others. Number down to 28 from 110 compared to method 1 and they have less deviation from the mean value.

- 2000-2002:

Companies satisfying the constraints are 5 out of 90 companies. Its includes Lexmark, First data corp, appl, amd, qlogic, sprint, nvidia, LSI and others. Number of companies down drastically from 42 to 6 when compared with method 1 and Only one company has some consistent performance. The computed values have less deviation from the mean value.

- 2003-2006:

  Companies satisfying the constraints are 15 out of 78. They include LEXMARK,MOBILE TELE, Seagate and others. Popular companies APPL, AMD, YAHOO, etc are removed which restricts us to a very few stocks for investment.

5.5 *Factor model and stock returns prediction*

5.5.1 Introduction

In this section stock returns are predicted using the factor model with key fundamental ratios as factors chosen based on the sector each stock belongs to. Specific numbers of stocks in each sector are chosen based on the methodology described in section 5.5. Factors for each stock to predict the future returns in a particular sector are determined using correlation analysis. Predicted returns are compared with the original returns to determine the accuracy of prediction. Structure of stock model, factors to include in the model, prediction methods and frequency of returns can be modified. Influence of macro, sector data, market rumors, sudden shocks, news, mergers, acquisitions, bankruptcy will be considered for future work.

5.5.2 Factor Model

Stock returns are based on the key fundamental ratios

$$y_{t+1} = c + \sum_{i=1}^{n} \beta_i F_t^i \quad (5.2)$$

Where

y_{t+1} – Stock Returns at time t+1

c - Regression constant

F_t^i - Fundamental ratio i at time t

β_i - Coefficient of the ratio i

5.5.3 Data

Stock fundamental data, is extracted from msn money where 10 years of historical data for key fundamental ratios are provided. Monthly data points are interpolated from the available yearly data. Monthly returns are also computed from Jan 2001 to Jan 2010 from historical closing price for each stock. Below are stocks selected in each sector. Stocks are selected on the universe reduction method discussed in section 5.4.2.1

- Sector: Basic

  Companies: Total SA, Exxon Mobile, British Petroleum, Chevron Corporation, Schlumberger, Arcelor Mittal, BHP Billiton Limited, Rio Tinto Plc

  Fundamental ratios: EPS, Price /Book ratio, Return on Assets, Price Sales ratio

- Sector: Conglomerates

  Companies: General Electric Co, United Technologies Corp, PPG Industries Inc, Textron Inc, Cooper Industries plc, Danaher Corp

Fundamental ratios: EPS, Price /Book ratio, Return on Assets, Debt to Equity ratio

- Sector: Consumer

  Companies: Toyota Motor Corp, Procter & Gamble Co, The Coca-Cola Company, Honda Motor Co, Pepsi co, British American Tobacco

  Fundamental ratios: EPS, Price /Book ratio, Price to Sales Ratio

- Sector: Financial

  Companies: Citigroup, Inc, JPMorgan Chase & Co, Morgan Stanley, American Express Company, The Goldman Sachs Group, U.S. Bancorp, Bank of America Corporation

  Fundamental ratios: EPS, Price /Book ratio, PE

- Sector: Health

  Companies: Pfizer Inc, Johnson & Johnson, Abbott Laboratories , Amgen Inc, GlaxoSmithKline, Medtronic, Inc, Novartis AG

  Fundamental ratios: EPS, Price /Book ratio, Return on Assets, Return on Equity

- Sector: Industrial

  Companies: Boeing Co, Tyco International Ltd, General Dynamics Corp, Northrop Grumman Corporation, Rockwell Automation Inc, Masco Corporation, Dover Corp, Vulcan Materials Company, Honeywell International Inc

  Fundamental ratios: EPS, Price /Book ratio, Return on Assets, Return on Equity, Debt to equity ratio, Net Profit margin

- Sector: Services

Companies: Wal-Mart Stores Inc, McDonald's Corp, Time Warner Inc, Lowe's Companies Inc, Amazon Inc, Kohl's Corp, Apollo group, Target Corp

Fundamental ratios: EPS, Price /Book ratio, Return on Assets, Net Profit Margin

- Sector: Technology

  Companies: AT &T Inc, Microsoft Corporation, Apple Inc, International Business Machines Corp, Intel Corporation, Oracle Corp, Hewlett-Packard Company, Verizon Communications Inc

  Fundamental ratios: EPS, Price /Book ratio, PE, Price to Sales Ratio

- Sector: Utilities

  Companies: Exelon Corp, Southern Company, Dominion Resources Inc, Waste Management Inc, Edison International, Consolidated Edison Inc

  Fundamental ratios: EPS, Price /Book ratio, Return on Assets, Return on Equity, Net profit margin

5.5.4 Implementation

- The data is split into training period and validation period.
- 11 months of fundamental data is used to fit the factor model.
- Key factors for each sector are used in correlation analysis
- Key factors having significant correlation with the month ahead stock returns for each stock are determined and are included as a part of the factor model.
- Coefficients / weights of the fundamental ratios are determined using recursive least square method.
- T+1 stock returns are predicted using the fundamental data at time t.

- New fundamental data is updated to the training window to improve the stock return predictions.
- Predicted data is compared with the original returns at the same time period to determine the accuracy.

5.5.5 Results

Mean error percent between the predicted stocks returns and original returns varies from stock to stock in each sector. Fig 16 and Fig 17 show the prediction error for stocks in consumer sector and utilities sector respectively.

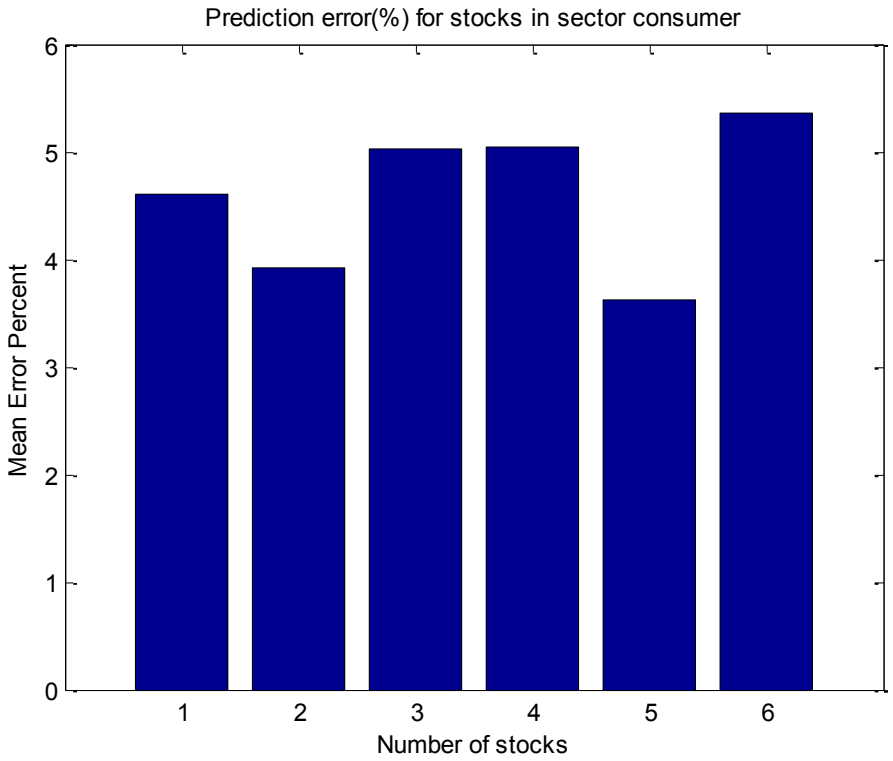

Fig 16: Prediction error (%) of stock returns in Consumer sector

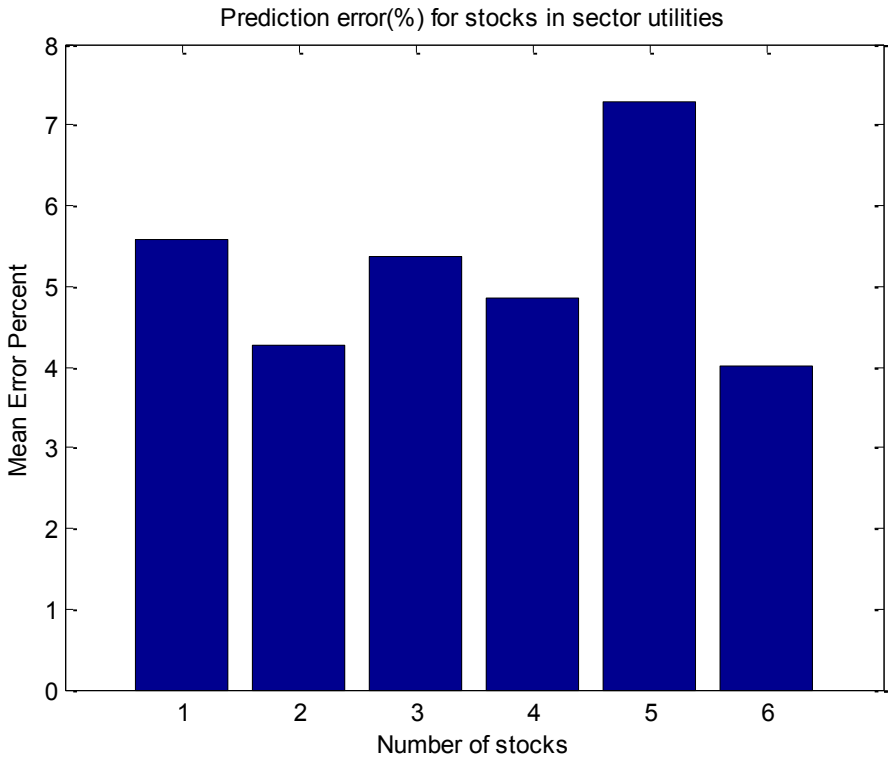

Fig 17: Prediction error (%) of stock returns in Utilities sector

5.5.6 Observations

Different methods to identify the right stocks based on their fundamental ratios were discussed. The entire data is divided into time periods based on the economic conditions and the fundamental values mean, minimum and maximum are computed with user defined constraints. Stocks obtained from the filtration process are used in the factor model to predict their future returns based on the key fundamental ratios identified. Correlation analysis is used to determine the important fundamental ratios which have the capability to predict the stock returns.

6 ASSET ALLOCATION/ OPTIMIZATION

6.1 *Introduction*

Wealth model is used for different portfolio optimization problems. Asset model described in chapter 4 is used in a receding horizon optimization framework. Advantages and drawback of different forms of risk are discussed and covariance risk is used as a risk constraint. Portfolio optimization methods such as constrained wealth maximization and constrained risk minimax are presented and compared. Covariance between the assets is considered as risk. Risk measure, investment horizon, risk constraints, reallocation frequency are few variables that can be changed. Shorting of assets and transaction costs will be considered for future work.

6.2 *Modern Portfolio Theory*

In 1952 Harry Markowitz developed a portfolio–selection technique which is called as modern Portfolio theory (Markowitz 1952). He emphasized on diversification of wealth among different assets/ securities to reduce the risk of the investment. Earlier investments were based on returns where the entire wealth is invested in maximum return yielding stock. Not until 1952 risk was given equal importance as return of a portfolio. This led to the powerful model showing the power of diversification which awarded the founder-Markowitz with a Nobel Prize in Economics.

The Markowitz efficient frontier consists of all possible portfolios who's returns are maximum at a given level of risk. Measurement of risk and return form a crucial part of the model. According to Markowitz two inputs were required for the model

- Expected return of every asset

- Variance of each asset and covariance between the assets used in the model.

The best portfolio is obtained by equation (6.1) or (6.2)

$$min\ (w^T \sum w)\ \text{s.t}\ \ R^T w \geq \mu \tag{6.1}$$

$$max\ (\ R^T w\ )\ \ \text{s.t}\ \ w^T \sum w < \sigma^2 \tag{6.2}$$

Where

w = Weights (allocation in each asset)

$\sum$ = Covariance Matrix

R = Average Returns (Historical or Analyst)

μ = Expected Portfolio Return

σ^2 = Expected variance of Portfolio

Variation in the expected return was considered as risk. Few other assumptions made by Markowitz are that investors would always want to maximize the expected future return with risk as low as possible. In-spite of the great work the model drew some criticism from many people. In the article *"The Markowitz optimization Enigma: Is "Optimized" Optimal?* (1989), Michaud discusses the practical problems of the model. Some of the problems are

- As there are no precise methods to compute the expected return, variances and the covariance matrix used in the model, the optimizer maximize errors. (Michaud 1989).

- The idea of using historical assets returns mean as the expected returns in the model is not a good method since history need not repeat itself.
- It was found that, small changes in the expected returns leads to drastic change in the portfolio allocation and the hence the model is unstable with respect to the returns.
- Markowitz model does not take into account the assets market capitalization weights. If there is an asset with low market capitalization and high returns then the model can suggest high portfolio weight to that asset.
- Typically Markowitz model suggest negative weights in assets indicating shorting when there are no constraints placed. When a constraint of no negative weights is placed on the optimizer results in zero portfolio weights to few assets and more weights to the rest. This results in concentration of funds in one or 2 assets.

6.2.1 Efficient frontier

The line along the upper edge of the region where every possible asset combination can be plotted in risk return space is called efficient frontier. All the portfolio combinations which lie on this line represent that portfolio which has the maximum return (y-axis) for the corresponding risk level represented on x-axis. Mathematically efficient frontier is the intersection of set of portfolios with minimum variance and the set of portfolios with maximum return (Wei-Peng Chen n.d.). We try to find out a portfolio which lies on the efficient frontier using the mean variance optimization.

6.3 *Risk Measures.*

6.3.1 Covariance:

Covariance matrices have widely been used for risk measurement in economic models and portfolio optimization (Primbs 2007) (Wolf 2003). This approach is used in this paper. The method of estimating covariance matrices often has significant impact on allocations (Winkelmann 1998). Recently work has been done on modifying/ estimating covariance matrices in noisy environments for mean-variance optimizations (Kondor 2002) (M.Wolf 2004).

6.3.2 Downside Deviation:

Citing the drawbacks of using covariance as a measure of risk in Mean-Variance optimization (MVO), downside risk optimization was proposed with downside risk as risk measure rather than standard deviation (Frank.A.Sortino & Price 1994).Post-Modern Portfolio Theory (PMPT) (Kasten 2005) explains the key reasons for using downside risk are, financial asset returns do not follow normal distribution, making the use of standard deviation inaccurate and any upside deviation from mean return or minimum acceptable return is always preferred and not viewed as risk by investor, which is not the case with MVO. (Kasten 2005) show that downside risk optimization is better mean variance optimization with numerical example.

6.3.3 Value at Risk – VAR:

Risk has been redefined in number of ways to overcome the problems which occur due to variance. One of the common risk management measures is Value at Risk (VaR). It is defined as the maximum potential loss in value of a portfolio for a given probability (ENGLE 2001). In simpler words, it's the amount which an investor can loose with a probability p over a given period of time. This was not popular until late 1990's even though Markowitz proposed

that considering downside variation of returns as risk leads to more efficient portfolios. (PHILIPPE ARTZNER 1999) shows that VaR is not a "coherent" measure of risk because it fails to satisfy the "sub-additive property." The VaR optimization problem is non-convex and may exhibit many local minima and results in exponential growth in computational complexity (Shapiro 2001), (U. S. Krokhmal 2001) and (Medova 1998) discuss the properties of VaR based-optimal portfolios acknowledging considerable computational difficulties. (Gaivoronski 2004) compares the efficient portfolios computed using covariance, VAR and CVaR as risk measure.

6.3.4 Conditional Value at Risk – CVaR:

This is an extension to the VaR, where the limitations of VaR such as the ability to limit the likelihood of incurring losses by certain types of risk and not by all risks are addressed. ie, the losses beyond the VaR are overlooked which might lead to substantial loss in portfolio value. It quantifies dangers beyond VaR. It also reduces the computational complexity leading to a numerical stability and efficiency of the calculations (Rockafellar 2001). (Gaivoronski 2004) Compares the efficient portfolios computed using covariance and CVaR as risk measure. (Jonas Palmquist1 1999) performs optimization where the returns are maximized with CVaR constraints.

6.4 *Wealth distribution among sectors*

Asset allocations are decided and wealth among the sectors is distributed with the goal of maximizing the returns with the minimal risk and is achieved using future returns and covariance's in the MVO (Mean variance optimization). Future expected returns and covariance among sectors are determined by

forecasting the returns and covariance using multiple regression model and hybrid model explained in chapter 5.

Even though Markowitz has said that historical returns and covariance alone are insufficient for estimating future returns and covariance, to make asset allocation decisions, we also realize from the literature the amount of information historical data contains which can be used to better decision making. One of the excerpts says

"…covariance matrices determined from empirical financial time series appear to contain such a high amount of noise that their structure can essentially be regarded as random. This seems, however, to be in contradiction with the fundamental role played by covariance matrices in finance, which constitute the pillars of modern investment theory and have also gained industry-wide applications in risk management "(Pafke 2002)

6.4.1 Implementations

Average daily return for a month and monthly return of the sector ETF's are predicted n steps into the future, using the sector asset model mentioned in 4.2.1. Monthly Covariance between sectors ETF's predicted values are computed. These forecasted values are used in Markowitz mean variance optimization with covariance risk constraint to obtain the portfolio weights.

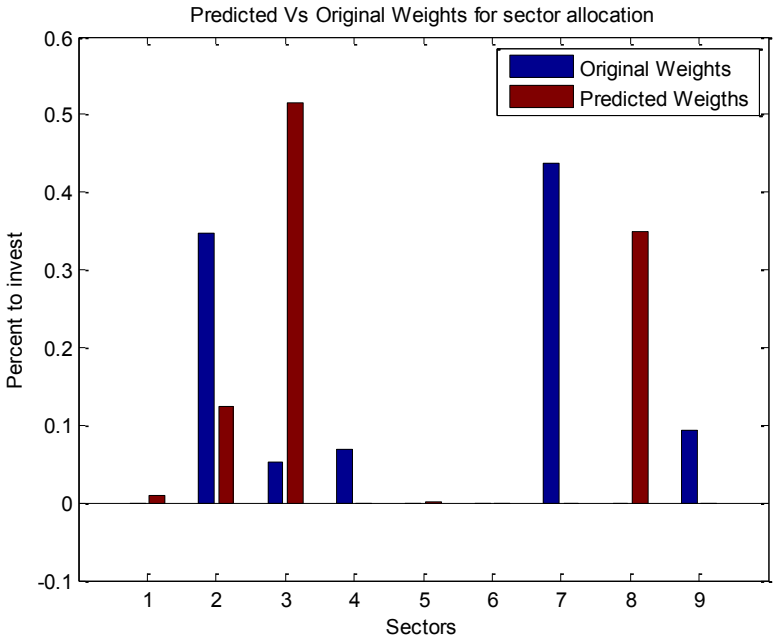

Fig 18: comparison between predicted and original allocation weights among sectors at time =k

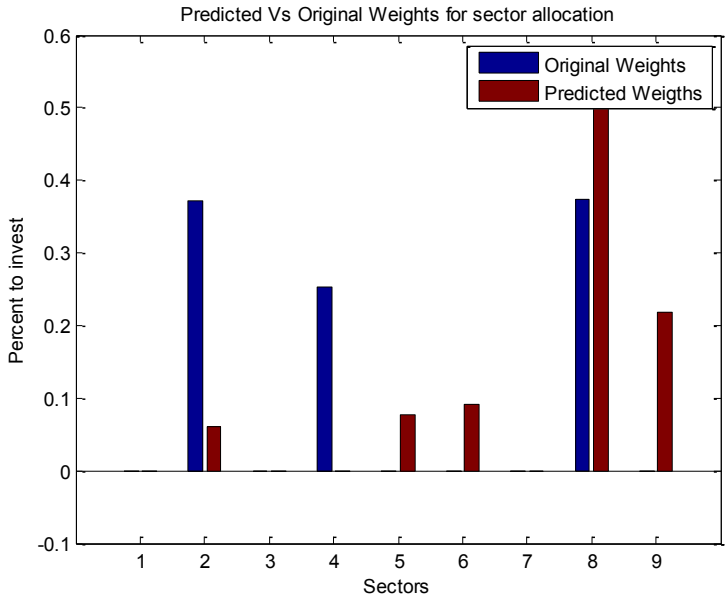

Fig 19: comparison between predicted and original allocation weights among sectors at time =k+1

Fig 20 shows the change in the portfolio return over a period of time, with the target monthly return 1.001 constraints and minimizing covariance risk as cost function in mean variance optimization.

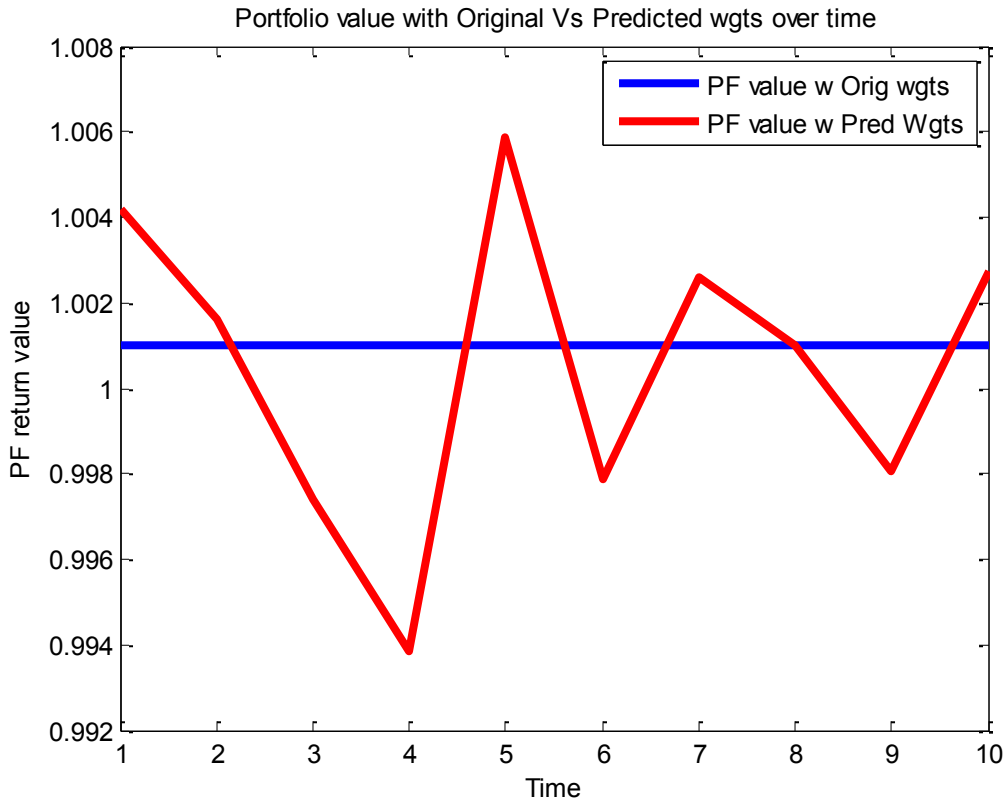

Fig 20: Portfolio returns trend using predicted returns forecast and covariance

6.4.2 Observations, Results and Problems

- As (Michaud 1989) rightly mentioned, the asset allocation greatly vary every month, due to the change in the expected return.
- Covariance matrix is highly ill-conditioned which results in significant variations in portfolio weights among the sectors.
- Average daily returns for a month has worst case prediction error of 3% when predicted 6 months ahead among all the sectors.

- Monthly return has higher worst case error.

The problem of matrix ill conditionality is tackled by flipping the objective of the problem. This is discussed in detail in section 6.5

6.4.2.1 Ill conditionality:

Condition number indicates the sensitivity of the output due to the small change in the input. Since the asset prices move with a greater frequency within a short period of time, the condition number is high. Magnitude of Variation in the asset prices reduces with long period, and hence decreasing the condition number. Due to the ill conditionality problem of the covariance matrix, monthly covariance prediction led to large variations in the asset allocation. Hence covariance matrix is computed using the entire training set to reduce the condition number.

6.5 *Portfolio Optimization with Receding Horizon Control*

The objective of any investor is to increase his/her investments over a period of time. Objective can be achieved by multiple strategies'. Identifying the weights of allocation between the assets with mean variance optimization and other methods, does not clearly convey the portfolio performance. Hence the same objective is achieved with different techniques which provide a greater insight in the wealth movement and portfolio performance for the entire investment period. Receding horizon control based portfolio optimization techniques use a wealth model placing constraints on the accepted risk and target wealth to maximize and track the target wealth. The wealth model is described below

6.5.1 Total Portfolio Wealth (Plant) Model

The total wealth associated with each asset at the start of period (k+1) is return on the total investment in the asset in period k. Given this, state space "plant" model for total portfolio wealth is represented by

$$\overline{\mathrm{w}}^{(\mathrm{k}+1)} = \mathrm{R}^{\mathrm{k}}\left(\overline{\mathrm{w}}^{\mathrm{k}} + \bar{\mathrm{u}}_{\mathrm{i}}^{\mathrm{k}}\right) \tag{6.3}$$

$$\mathrm{W}^{\mathrm{k}} = [1\ 1\ 1\ \ldots]\overline{\mathrm{w}}^{\mathrm{k}} \tag{6.4}$$

Where,

- $\overline{w}^k$: Total Wealth vector at start of period k
- R^k : Diagonal matrix of returns of $\overline{r}_i^k$
- $\bar{u}_i^k$: Reallocation vector at the start of period k
- $\overline{w}^k + \bar{u}_i^k$: Total capital vector invested during period k
- W^k : Total portfolio wealth at the start of period k

The initial conditions (at k=1) for the wealth is w ≡ 0. $\overline{w}^k$ is considered the plant's state vector, $\bar{u}_i^k$ represents the plant's control vector and W^k the plant's output.

6.5.2 Total Portfolio Wealth (Plant) Model Constraints

$$\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{u}_{\mathrm{i}}^{\mathrm{k}} = \mathrm{q}^{\mathrm{k}} \tag{6.5}$$

$$\mathrm{u}_{\mathrm{i}}^{\mathrm{k}} \geq -\mathrm{w}_{\mathrm{i}}^{\mathrm{k}} \tag{6.6}$$

Where,

- q^k : Cash flow at start of period k.

6.5.3 Total Portfolio Wealth (Plant) Model Assumptions

The following assumptions are inherent in our total portfolio wealth (plant) model:

1. The return entries along the diagonal of R are estimated as discussed in section 4.2;
2. The constraint in equation 6.5 requires that the total reallocation from different assets at the start of a period must equal the cash flow for that period; i.e. all available wealth must be invested. However a risk free asset is included in the model to represent wealth that is not invested in any other asset, and it has a unit rate of return.
3. The constraint in equation 6.6 precludes selling more of an asset than what's present in the portfolio. As such, it prevents shorting of assets. In this simple model, any short would have unlimited downside potential.
4. The cash flow is known *a priori* for the entire control horizon.

6.5.4 Constrained wealth maximization

In this problem, wealth is maximized subject to risk tolerance constraints i.e. the risk in each period (as measured by the covariance of the daily returns series) to be less than a specified tolerance. The optimization problem is as described below,

$$\max_{\mathrm{u}} || \bar{\mathrm{w}}^{\mathrm{k+h}} ||_1 \qquad (6.7)$$

Where $|| v ||_1 = \sum_i |v_i|$ is the vector one norm of the vector $v = [\, v_1, \ldots]^T$.

6.5.4.1 Classic Risk Constraint

In addition to constraints (6.5) and (6.6) the following constraint is imposed

$$\left(\bar{w}^{i}\right)^{T}\sum\bar{w}^{i} \leq \sigma_{r} \max\left(\mathrm{diag}(\sum)\right) \ \forall\, i \in \{k, k+1, \ldots, k+h\} \qquad (6.8)$$

where

- k: Starting period of the prediction horizon
- h: Number of periods in the prediction horizon
- $\sum$: Covariance matrix of daily returns
- σ_{r}: Constant to adjust risk tolerance ($\sigma_{r} \in (0, \infty)$)
- $\bar{w}^{i}$: wealth vector for the ith period

$\sum_{i,j} = \mathrm{cov}(\bar{r}_{i}, \bar{r}_{j})$ where cov(s, t) is the covariance between the two time series s and t, and $\bar{r}_{n}$ denotes the return time series for the asset n. σ_r is a constant (risk decision parameter) used to adjust risk tolerance ($\sigma_r \in [0, \infty)$). $\sigma_r = 0$ represents no risk (i.e. infinitely conservative). $\sigma_r = \infty$ represents infinite risk (i.e. infinitely aggressive). $||\sum||_2$ is the largest singular value of $\sum$. Since $\sum^T = \sum$, $||\sum||_2$ is also the largest eigenvalue of $\sum$.

Increasing (decreasing) σ_r corresponds to an investor with higher (lower) risk tolerance. Since a risk free asset is included, there is a feasible solution for all positive σ_r. $\sigma_r \geq 1$ would imply the investor is not risk-conscious (arbitrarily high risk tolerance). h is chosen based on the accuracy of returns prediction for the desired horizon. For calculating $\sum$, all available data up to

the current period k is used. The condition number of the covariance matrix increases as the length of data used for its computation reduces.

Equation (6.8) places an upper bound on the risk for every period of the prediction horizon. Risk constraints are only placed on the allocation at the start of the period; however, due to the low condition number of, its believed that changes in wealth in each asset during the period (as certain returns rise and fall) does not significantly affect the risk on the overall portfolio.

6.5.5 Constrained risk minimax

There are several ways of estimating an investor's risk tolerance (Luenberger 1998). Typically, investors would be unable to specify the risk-aversion parameter σ_r used in equation (6.8). Certain risk measures, such as VaR, are more accessible to investors than others. Significant work has been done in trying to estimate the investor's preference among risky assets (Grable 2001) (M. J. Roszkowski 2005) (M. S. Kimball 2008). However, investors would be better able to specify their expected/minimum acceptable returns. A wealth target (known a priori) is considered for each period of the horizon. The objective is to minimize the peak risk (over all periods in the prediction horizon) subject to constraints on the wealth in each period. A similar problem can be found in (G. D. F.Herzog 2007), where a trade-off between risk and asset return is considered. In (Yang 2001) the authors attempt to minimize the average peak individual risk of the stocks; in (X. Cai 2000) (Young 1998), the objective is to minimize 'the expected absolute deviation of

future returns from their means'. (Young 1998) (X. Deng 2005) considers the problem of maximizing the worst-case expected returns of the portfolio.

6.5.6 Minimax Risk Objective

Risk objective is defined as

$$\min_{u} \; || \left(\bar{w}^{i}\right)^{T} \sum \bar{w}^{i} \;\; , i \in \{k, k+1, \ldots, k+h\}^{i} \; ||_{\infty} \qquad (6.9)$$

Where

$||.||_{\infty}$ denotes the peak value of $(\bar{w}^{i})^{T} \sum \bar{w}^{i}$ over the time window $i \in \{k, k+1,\ldots,k+h$

6.5.7 Total Wealth Constraint

The following wealth constraint is defined (in addition to constraints (4) and (5))

$$|| \bar{w}^{i} ||_{1} \geq \tilde{w}^{i} \; \forall \, i \in \{k, k+1, \ldots, k+h\} \qquad (6.10)$$

where,

- $\tilde{w}^{i}$: Target wealth for period i

Depending on $\tilde{w}^{i}$ and the expected returns, there might not exist a feasible solution to the constraint ie the wealth constraint may be too aggressive. In such a scenario, the wealth is invested in the asset with maximum expected returns (equivalent to σ_{r} = 1 in the constrained wealth maximization scenario), as this allocation brings the (expected) portfolio wealth 'closest'to $\tilde{w}^{i}$ investor wealth target.

6.5.8 Implementation

Both problems discussed in section 6.5.4 and 6.5.5 using two ETF classes - Standard & Poor's Depositary Receipt (SPDR - benchmark: S&P 500), and iShare (benchmark: Dow Jones Industrial Average) are simulated. The test period is January 2006 to January 2010. The benchmarks rise and fall in this period, helps in observing the algorithm's performance in diverse economic environments. A cash flow of $100 is assumed to be available at the start of the investment period. There is no other external cash flow to/from the portfolio. Transaction costs have also been neglected. A five-month prediction horizon is used.

6.5.9 Data

Macroeconomic data series are obtained from the National Bureau of Economic Research and the Federal Reserve Bank of St. Louis. GDP, Consumer Price Index, Ten-Year Treasury rate, Prime loan rate, Outstanding Consumer Credit, Industrial Production Index, Money Supply and Unemployment rate are obtained from January 1959 to January 2010. Sector ETF data are obtained from Yahoo! Finance.

6.5.9.1 Parameters

- q: Cash flow - $100 inflow in the first period and no transactions in the following periods
- h: Chosen based on reliable prediction horizon (5 months) Transactions costs were neglected
- $\sum$: Covariance matrix using all available data
- σ_r: Risk adjustment for method 1 - family of designs considered

- $\widetilde{w}^i$: Target wealth for method 2 - family of designs considered

6.5.10 Results

6.5.10.1 Constrained wealth maximization.

Results of method 1 are considered. Plots with results for various values of risk-tolerance σ_r appearing in equation {6.8} are shown. The result of the RHC can be seen in Fig 21 and Fig 22 (wealth of portfolio), and Fig 23 - Fig 26(controls). RF denotes a risk free asset (i.e. an asset with unit returns). The benchmark portfolio performance is shown as a dotted line. Table 21 summarizes the results from Fig 21 and Fig 22.

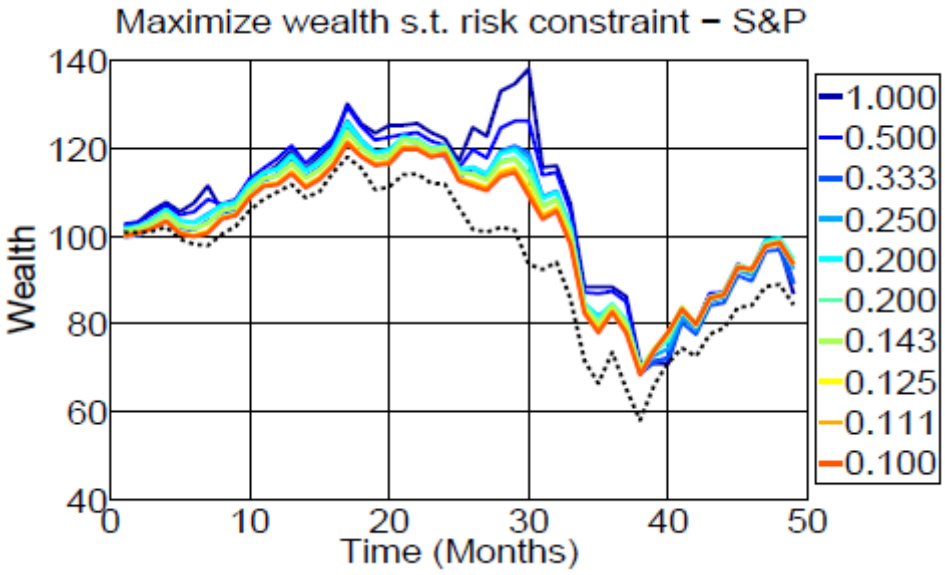

Fig 21 : SPDR- Total Wealth

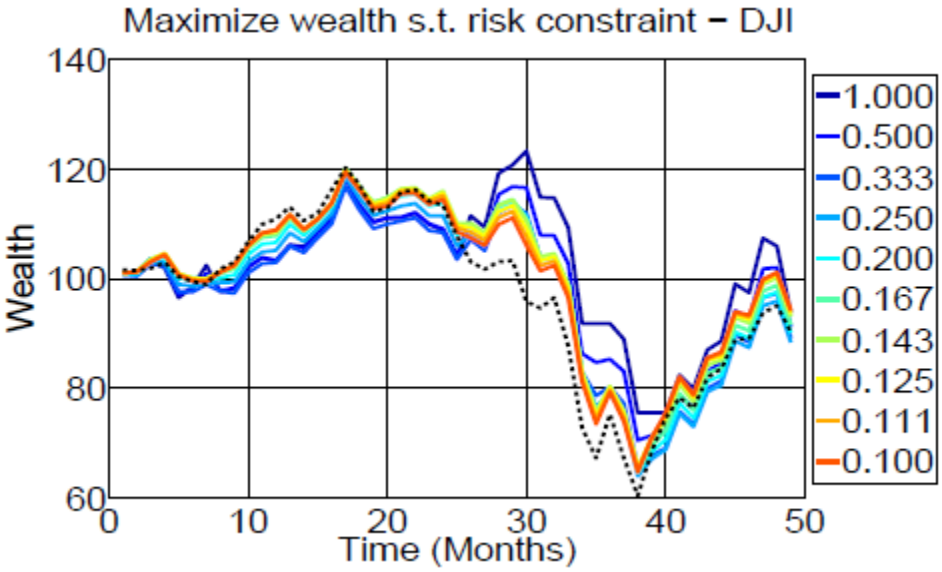

Fig 22 : iShare – Total Wealth

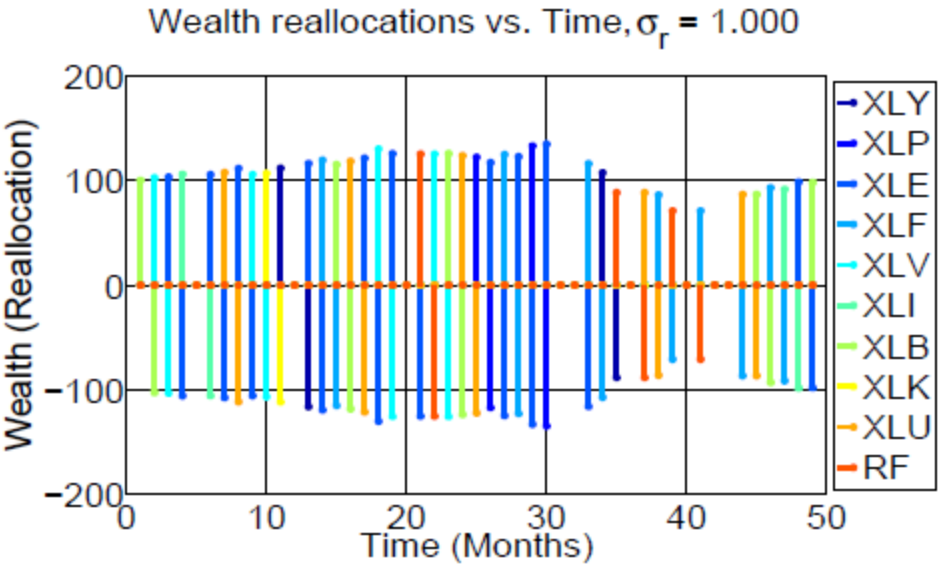

Fig 23: SPDR – Controls for $\sigma_r = 1$

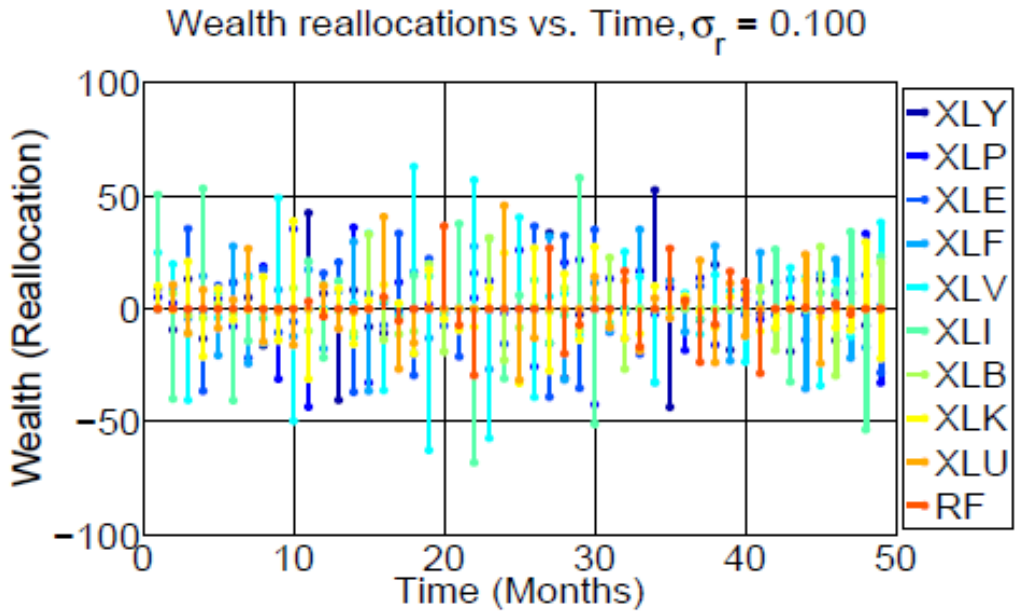

Fig 24 : SPDR – Controls for $\boldsymbol{\sigma_r = 0.1}$

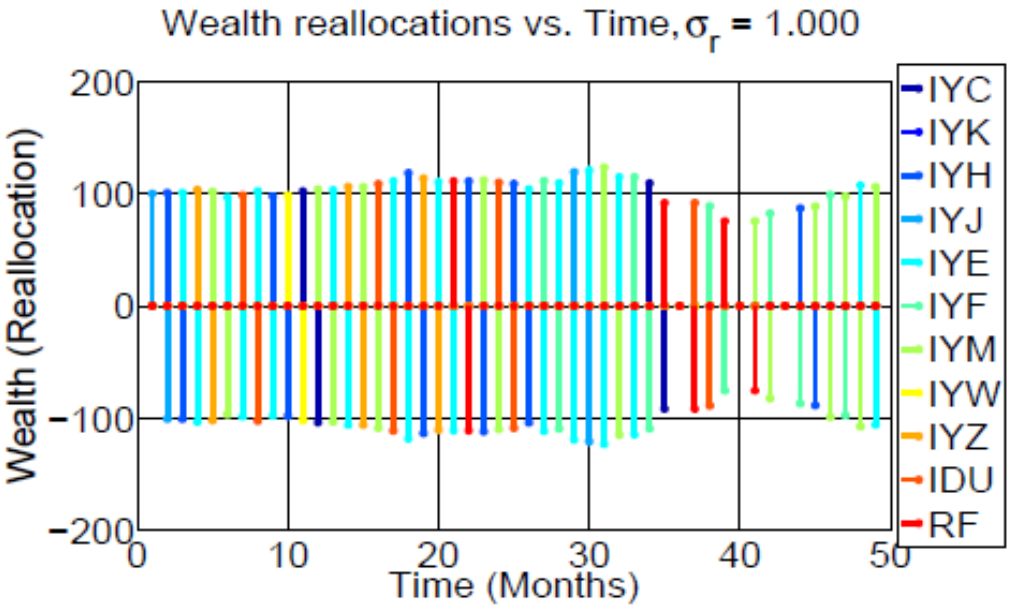

Fig 25: iShare – Controls for $\boldsymbol{\sigma_r = 1}$

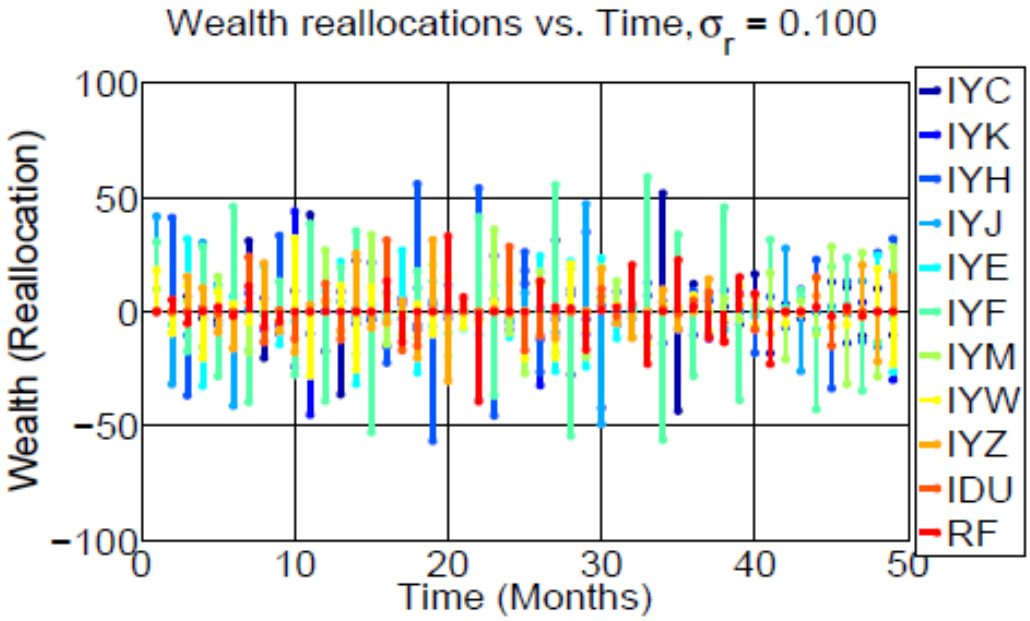

Fig 26: iShare – Controls for $\boldsymbol{\sigma_r = 0.1}$

From Fig 23, Fig 24, Fig 25 and Fig 26 the following is observed

- Without risk constraints ($\tau \leq 1$): In each period, all the wealth is concentrated in a single (maximum expected return) asset
- As σ_r increases, diversification increases. However as σ_r is decreased further, the allocations might be concentrated in less risky assets.

From Fig 21 and Fig 22, the following is observed:

- In a bear market, low risk strategies outperform high risk strategies (less loss).
- In a bull market, high risk strategies outperform low risk strategies (more gains).

Table 21: Wealth Maximization

σ_r	SPDR	iShare
1 (Risky)	86.77	93.69
0.5	89.22	91.14
0.33	89.12	89.12
0.24	92.45	88.34
0.2	94.79	90.1
0.16	94.03	91.62
0.14	93.94	93.03
0.12	93.33	93.61
0.11	93.51	94
0.1 (LowRisk)	93.45	94.26
Benchmark	84.3	90.29

6.5.10.2 Constrained risk minimax.

The problem of minimizing peak risk over the prediction horizon subject to a wealth tracking constraint is considered. It has two alternate approaches:

- Fixed target: In this scenario, the wealth target is fixed a priori for the whole investment period.
- Variable target: In this case, the target is set based on the desired return (from the current state). However, if the benchmark outperforms the desired return, the same gains over the current benchmark is targeted (i.e. Target is never less than the benchmark performance).

In the second approach, each iteration resets the target based on the current state. Hence growth is attempted, without taking too much risk. In the case of fixed target, the following scenarios merit further analysis:

- Under-performance: If the performance is much below target, a risky strategy with high growth is attempted (even in a bear market).

- Over-performance: If the performance is greater the target, the optimizer does not attempt any growth (even in a bull market).

If there does not exist a solution satisfying the wealth tracking constraint {6.7}, both strategies invest all the wealth in the asset with the maximum expected return. Fig 27 and Fig 28 show the wealth of this approach for SPDR ETF's. Table 22 summarizes the results of this methodology

Table 22: Risk Minimax

Desired Monthly Gain	SPDR		iShare	
	Fixed	Variable	Fixed	Variable
0.10%	103.79	112.75	104.4	110.69
0.50%	96.61	106.74	96.49	101.08
1.00%	86.2	97.97	94.73	87.17
1.50%	84.83	89.16	92.28	86.27
2.00%	84.22	87.07	92.78	85.93
5.00%	86.77	83.23	93.69	90.23
10.00%	86.77	80.4	93.69	93.03
Benchmark	84.3		90.29	

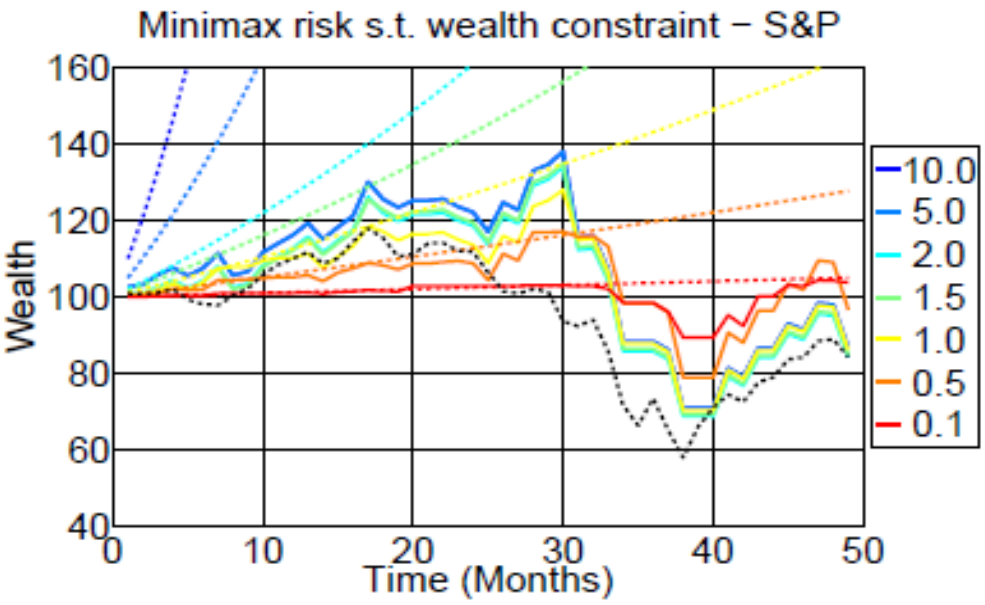

Fig 27 : SPDR Total Wealth – Fixed Target

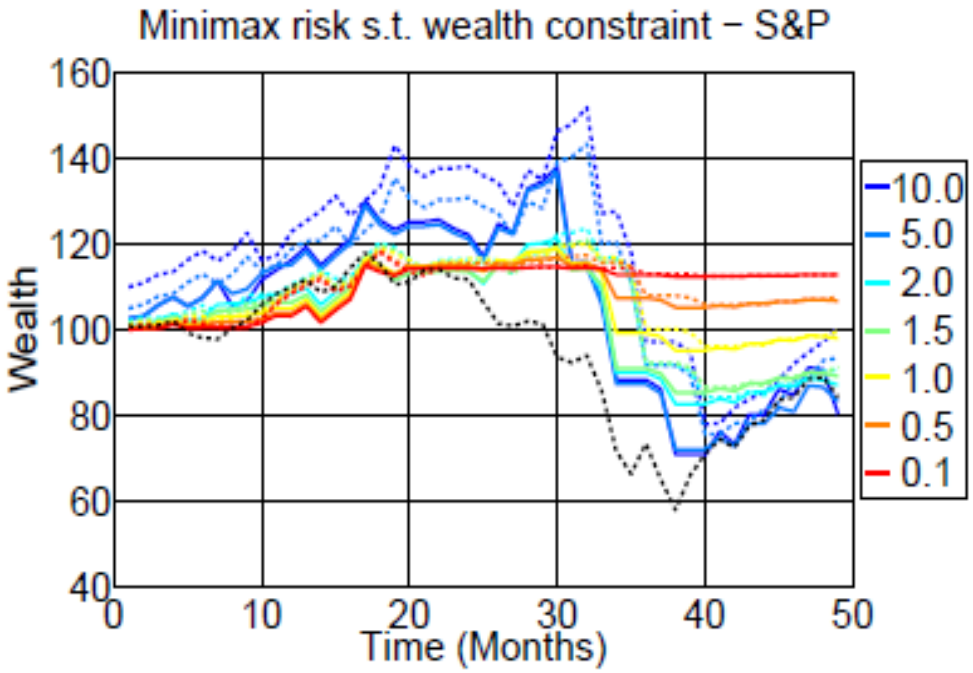

Fig 28 : SPDR – Total Wealth – Variable Target

- When high returns are demanded, the result is similar to wealth maximization with a high risk tolerance.
- During and after a bear market, a low returns-variable target approach (LR-VT) performs better than other risk minimizations, method 1 and the benchmark

6.5.11 Comparisons

In a bull market, CWM with high risk tolerance performs similar to CRM with high expected returns. All risk-tolerances of CWM lose significantly in bear markets, while a LR-VT strategy with CRM retains much of its earlier gains. Since the accurate timing of business cycles is a difficult (open) problem, such a low-return strategy using CRM can prove useful. If it was possible to time the business cycle accurately, both methods could be adjusted to make use of this additional information. During the growth phase, both of them would perform similarly. However, in a bear market, CRM would still retain most of its gains.

6.6 *Constrained risk minimization using stocks and ETF's*

Wealth allocation among stocks and sector ETF's is explored. Specific stocks in each sector are selected using stock selection methods discussed in chapter 5. Stock returns are predicted using stock model. Constrained wealth maximization (CRM) and constrained risk minimization with fixed and variable target are used. Wealth allocated in sector (as discussed in section 6.5) is distributed among selected individual stocks by minimizing the risk in each period (as measured by the covariance of the daily returns series) subject to return constraints i.e. stock returns to be greater than or equal to sector ETF returns.

6.6.1 Optimization problem

The optimization problem is as described below,

$$\mathbf{min} \quad ||(\bar{\mathbf{w}}^{\mathbf{i}})^{\mathbf{T}} \sum \bar{\mathbf{w}}^{\mathbf{i}}||_2 \qquad (\mathbf{6.11})$$

Where

- $\sum$: Covariance matrix of daily returns
- $\bar{w}^i$: fractional wealth vector for the ith period

6.6.1.1 Return Constraints

$$R^T \bar{\mathrm{w}}^{\mathrm{i}} \geq \mu \qquad (6.12)$$

Where

- R : Stock returns in each sector
- $\bar{w}^i$: fractional wealth vector for the ith period
- μ : Sector ETF returns

For calculating $\sum$, all available daily data up to the current period k is used. The condition number of the covariance matrix increases as the length of data used for its computation reduces.

6.6.2 Results

Two different optimization methods are used with stocks and sector ETF's as assets. In both the methods, wealth allocated for a particular sector is distributed among the stocks within that sector if the constraint has a feasible solution. When the constraint is not feasible

Case 1 (only Stocks): Wealth is allocated to highest return stock.

Case 2 (Stocks + Sector ETF's): Wealth is allocated to sector ETF.

6.6.2.1 CWM:

Investing in stocks + sector ETF's (Fig 30) is better than investing only in stocks (Fig 29). ETF's are assumed to posses diversification property and hence its advantageous to include sector ETF's in portfolio.

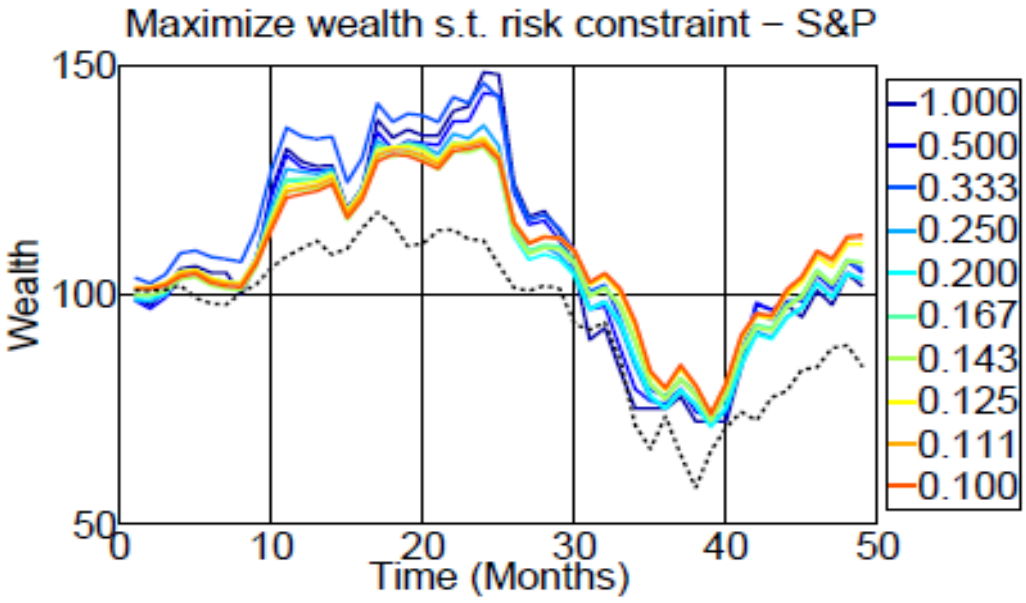

Fig 29: CWM- Wealth allocated among Stocks

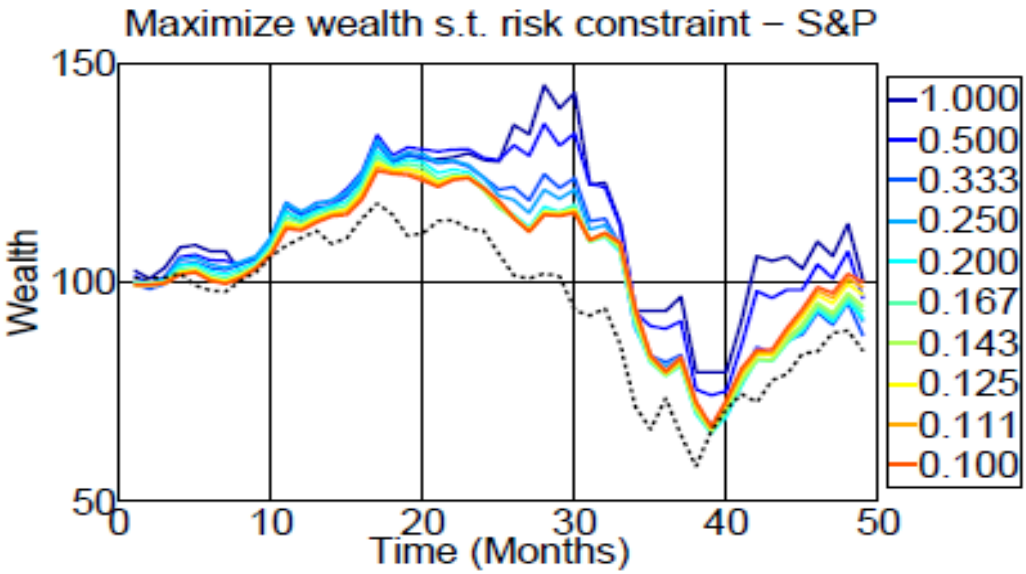

Fig 30: CWM- Wealth allocated among Stocks+ Sector ETF's

6.6.2.2 CRM – Fixed target

In both the scenarios low targets are achieved compared to high targets.

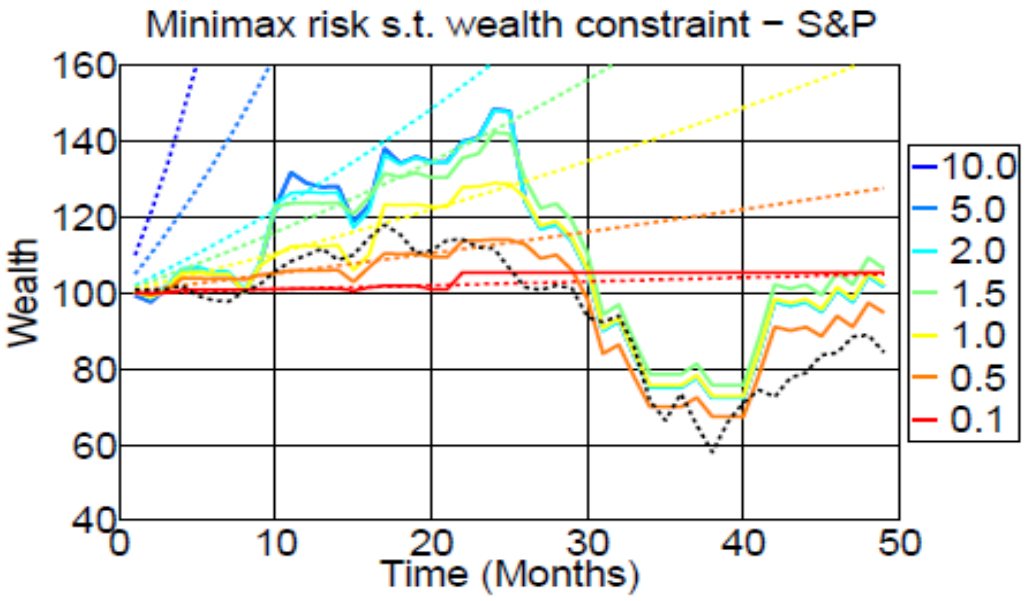

Fig 31: CRM - Fixed Target with stocks

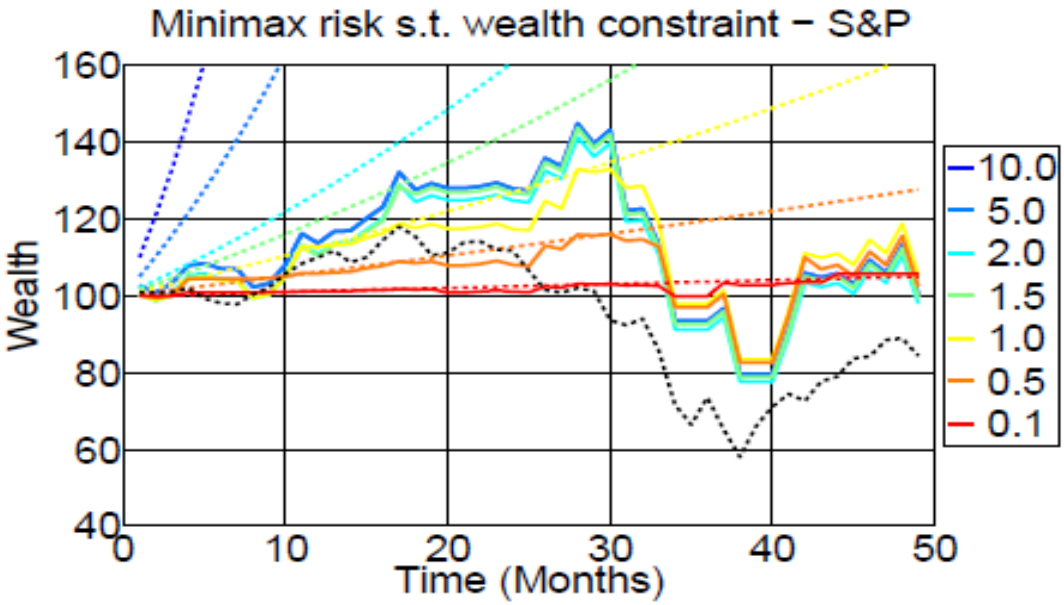

Fig 32: CRM - Fixed Target with stocks + sector ETF's

6.6.2.3 CRM – Variable Target

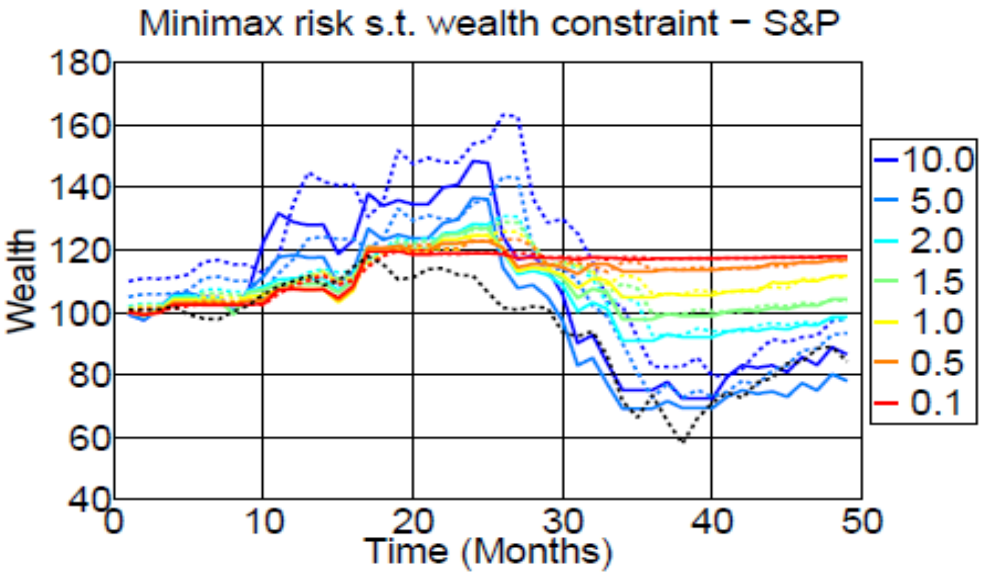

Fig 33: CRM - Variable target with stocks

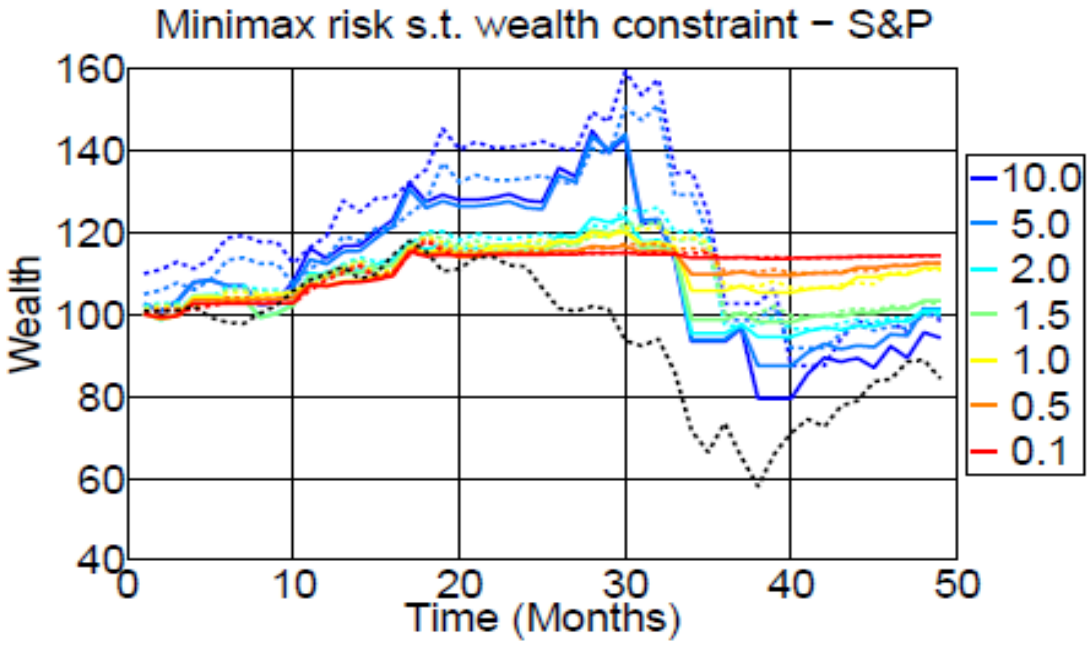

Fig 34: CRM- Variable target with stocks+sector ETF's

6.6.2.4 Comparisons:

In bull market, CWM (Fig 29, Fig 30) results in sharp rise in gains when stocks are considered for investment. CRM- fixed target method can achieve reasonable targets (less than 2%) when stocks (Fig 31) or combination of stocks and ETF's (Fig 32) are considered. CRM-variable target method (Fig 33, Fig 34) tracks the desired target almost exactly when the expected returns are less than or equal to 2%.

In bear markets, CWM and CRM-fixed method indicates that stocks+ sector ETF's perform much better with high gains compared to the portfolio with only stocks. In CRM-variable target method, high return target (>5%) looses significantly than the low return target with only stocks in the portfolio. However when sector ETF's are included to the portfolio high return targets perform better than the benchmark returns. Low return targets are achieved without losing the portfolio value.

7 SUMMARY AND DIRECTIONS FOR FUTURE RESEARCH

7.1 *Summary*

In this thesis, a modular framework (environment) is put in place which enables systematic approach for data acquisition, portfolio components (macro, sector, stocks) analysis, modeling, prediction and management (decision making). Reasonable approach has been adopted to include macro, sector and stock data for analysis, modeling, prediction, functional dependence and decision making. Different models, methodologies and techniques have been explored to meet common investor's objectives. State-space factor and wealth model was used to address the problem of portfolio optimization. Macro-economic factors were central to the prediction of asset returns. Using ETFs as the risky assets, returns in excess of the benchmark portfolio were obtained. Stocks within each sector are identified using stock selection methods. As an alternative to wealth maximization to risk constraints, the problem of wealth tracking subject to risk minimization has been considered. This methodology performed better than the wealth maximization approach.

7.2 *Future Work*

Future work will include

- Identification of Business cycle dates earlier than NBER to improve accuracy of macro predictions.
- Studying more sophisticated algorithm for macro, asset, stock returns prediction.

- Develop complex algorithms for strategy switching between high risk /low risk as per bull and bear times.
- Develop machine learning algorithms for stock grouping and selection.
- Real time data collection and analysis software tool to be developed (MATLAB based)

REFERENCES

1. A.Protopapadakis, Mark J Flannery and Aris. "Macroeconomic factors do influence Aggregate stock returns." *Review of Financial studies*, 2002: Vol 15, no 3, 751-782.

2. Anne Sophie E Vanroyen, Sebastien Page. "The Multiple Dimensions of Asset Allocation: Countries, Sectors or Factors?" *Working Paper no 6.* State street Associates, 2002.

3. B. Durtschi, M. Skinner, and S. Warnick. ""Portfolio optimization as a learning platform for control education and research." *American control Conference*, 2009: 3781-3786.

4. B. I. Jacobs, K. N. Levy, and H. M. Markowitz. "Portfolio optimization with factors, scenarios, and realistic short positions." *Operations Research*, 2005: 586 – 599.

5. Batchelor, G. T. Albanis and R. A. *Five classification algorithms to predict high performance stocks.* Springer-Verlag, 2000.

6. Bertsekas, D. P. "Necessary and sufficient conditions for existence of an optimal." *Journal of Economic Theory*, 1974: 235-247.

7. Bertsekas, D. P. *Dynamic Programming & Optimal Control.* Athena Scientific, 2005.

8. Best, M. J. "On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results." *The Review of Financial Studies*, 1991: 315-342.

9. Bodie, Z. "Common Stocks as a hedge against inflation." *Journal of Finance*, 1976: 459-470.

10. Bollerslev, T., Cai, J. and Song, F.M. "Intraday Periodicity, Long Memory Volatility, and Macroeconomic Announcement Effects in the U.S.Treasury Bond Market." *Journal of Empirical Finance*, 2000: 37-55.

11. Box, Jenkins,Reinsel. *Time series analysis: forecasting and control.* Old Tappan,NT: Prentice Hall, 1994.

12. Boyd, Jagannathan,hu. *The Stock markets reaction to unemployment news: Why Bad News is Usually Good for Stocks.* Working Paper 8092, NBER, 2001.

13. Buckland, Maurice G Kendall and William R. *A Dictionary of Statistical Terms.* New York: Hafner Publishing Company, 1971.

14. *Bureau of Economic Analysis.* http://www.bea.gov/newsreleases/national/gdp/gdpnewsrelease.htm.

15. *Bureau of labour statistics.* http://www.bls.gov/bls/inflation.htm.

16. Carlos M. Gutierrez, Cynthia A. Glassman,J. Steven Landefeld. "Measuring the Economy: A Primer on GDP and the National Income and Product Accounts." Washington: Bureau of Economic Analysis (BEA), September 2007.

17. Castanias, RP. "Macroinformation and the variability of stock market prices." *Journal of Finance*, 1979: 439-450.

18. Chen, A Ross,Roll. "Economic Forces and the Stock Market." *The Journal of Business*, 1986: 383-403.

19. Chen, Z. "Numerical methods for optimal stochastic control in finance." Ph.D. dissertation, University of Waterloo, 2008.

20. Damodar N Gujarati, Sangeetha. *Basic Econometrics.* New York: Tata McGraw-Hill, 2007.

21. Deng, Y. *Genetic algorithms for financial portfolio selection.* Masters Thesis, Beijing: University of Science and Technology Beijing, 2002.

22. Doege, H. Luthi and J. "Convex risk measures for portfolio optimization and concepts of flexibility." *Mathematical Programming*, 2005: 541-559.

23. EF Fama, R Kenneth. "Dividend yields and expected stock returns." *Journal of Financial Economics*, 1988: 3-25.

24. ENGLE, SIMONE MANGANELLI and ROBERT F. *VALUE AT RISK MODELS in Finance.* Working Paper No 75, Germany: European Central Bank, 2001.

25. F. Herzog, S. Keel, G. Dondi, L. M. Schumani, and H. P. Geering. "Model predictive control for portfolio selection." *American Control Conference*, 2006: 1252-1259.

26. F.Herzog, Dondi, S. G Keel, L. M. Schumani, and H. P. Geering. "Solving alm problems via sequential stochastic programming." *Journal of Quantitative Finance*, 2007: 231-244.

27. F.Herzog, G. Dondi, and H. P. Geering. "Stochastic model predictive." *International Journal of Theoretical and Applied Finance*, 2007: 203-233.

28. Fama, Eugene F. "Stock Returns, Real Activity, Inflation and Money ." *American Economic Review*, 1981: 545-565.

29. Fleming, M. J. and Remolona, E. M. "What Moves the Bond Market." *Economic Policy Review*, 1997: 31-50.

30. Frank K. Reilly, Keith C. Brown. *Investment Analysis and Portfolio Management .* South-Western College, 2002.

31. Frank.A.Sortino & Price, L. N. "Performance Measurement in a Downside Risk Framework." *The journal of investing*, 1994: 59 - 64.

32. French, Eugene F. Fama and Kenneth R. "The Cross-Section of Expected Stock Returns." *The Journal of Finance*, 1992: 427-465.

33. Gaivoronski, A. A. & Pflug, G. "Value-at-Risk in Portfolio Optimization: Properties and Computational Approach." *Journal of Risk*, 2004: 1 - 31.

34. Grable, M. J. Roszkowski and J. "Estimating risk tolerance: The degree of accuracy and the paramorphic representations of the estimate." *Financial Councelling and Planning*, 2001: 29-41.

35. Green, C.T. "Economic News and the Impact of Trading on Bond Prices." *Journal of Finance*, 2004: 59, 1201- 1233.

36. H. Peyrl, F. Herzog, and H. P. Geering. "Numerical solution of the hamilton-jacobi-bellman equation for stochastic optimal control problems." *WSEAS International Conference on Dynamical Systems and Control.* 2005.

37. Hazlitt, Henry. *what should you know about inflation.* New Jersey: D. Van nstrand Company, Inc.

38. Investopedia. http://www.investopedia.com/terms/m/moneysupply.asp.

39. —. *Investopedia.* http://www.investopedia.com/terms/a/autocorrelation.asp (accessed 08 15, 2010).

40. —. *Investopedia.* http://www.investopedia.com/terms/e/equity.asp.

41. Irving B. Kravis, Alan W. Heston and Robert Summers. "Real GDP Per Capita for More Than One Hundred Countries." *The Economic Journal*, 1978: 215-242.

42. Jalali, A.A and V. Nadimi. "survey on robust model predictive control from 1999-2006." *International Conference on Intelligent Agents, Web Technologies and Internet Commerce.* 2006. 207.

43. Janice Peterson, Margaret Lewis. *The Elgar Companion to feminist economics.* Massachusetts: Edward Elgar Publishing, Inc, 1999.

44. Jeffrey Jaffe, Donald B. Keim and Randolph Westerfield. "Earnings Yields, Market Values, and Stock Returns." *The Journal of Finance*, 1989: 135-148 .

45. Jefrey Pontiff, Lawrence D. Schall. "Book-to-market ratios as predictors of market returns." *Journal of Financial Economics*, 1998: 141 - 160.

46. Jonas Palmquist1, Stanislav Uryasev2 & Krokhmal, P. *Portfolio Optimization With Conditional Value-at-risk Objective And Constraints.* Tech report, University of Florida: Center for Applied Optimization, 1999.

47. Kasten, P. S. G. W. "Post-Modern Portfolio Theory." *Journal of Financial Planning*, 2005: 74.

48. King, Benjamin F. "Market and Industry Factors in Stock Price Behavior." *The Journal of Business*, 1966: 139-190.

49. Kondor, S. Pafka and I. "Noisy covariance matrices and portfolio optimization." *The European Physical Journal B*, 2002: 277-280.

50. Lamont, OA. "Economic Tracking Portfolios." *Journal of Econometrics*, 2001: 161-184.

51. Lamont, Owen. "Earnings and Expected Returns." *The Journal of Finance*, 1998: 1563–1587.

52. Lansing, Kevin J. *Growth in the Post-Bubble Economy.* Newsletter, San Francisco: Federal Reserve Bank of San Francisco, 2003.

53. lehman, paige. *what is institutional ownership.* 04 2004. http://e-articles.info/e/a/title/What-is-Institutional-Ownership/.

54. Lewelle, Jonathan. "The time-series relations among expected return, risk, and book-to-market*1." *Journal of Financial Economics*, 1999: 5-43.

55. Livingston, Miles. "Industry Movements of Common Stocks." *The Journal of Finance*, 1977: 861 -- 874.

56. Lofberg, J. *Minimax approach to robust model predictive control.* Ph.D. dissertation, Linkoping University, 2003.

57. Luenberger, D. G. *Investment Science.* Oxford University Press, 1998.

58. M. J. Roszkowski, G. Davey, and J. E. Grable. "Insights from psychology and psychometrics on measuring risk tolerance." *Journal of financial planning*, 2005: 66-75.

59. M. S. Kimball, C. R. Sahm, and M. D. Shapiro. "Imputing risk tolerance from survey responses." *Journal of the American Statistical Association*, 2008: 1028-1038.

60. M. V. Kothare, V. Balakrishnan, and M. Morari. "Robust constrained model predictive control using linear matrix inequalities." *Automatica*, 1996: 1361 – 1379.

61. M.Wolf, O. Ledoit. "Honey i shrunk the sample covariance matrix." *Journal of Portfolio Management*, 2004: 110-120.

62. Makridakis, A. S., Andersen, R. Carbone, R. Fildes, M. Hibon,R. Lewandowski, J. Newton, E. Parzen, and R. Winkler. "The accuracy of extrapolation (time series) methods: Results of a forecasting competition." *Journal of Forecasting*, 1982: 111 – 153.

63. Markowitz, Harry. "Portfolio Selection." *The Journal of Finance*, 1952: 77-91.

64. McKenzie, E. S. Gardner and E. "Forecasting trends in time series." *Management Science*, 1985: 1237 – 1246.

65. Medova, E. "Var methodology and the limitation of catastrophic or." 1998.

66. Merton, R. C. "Lifetime portfolio selection under uncertainty: The." *Review of Economics and Statistics*, 1969: 247 – 267.

67. Michaud, Richard O. "The Markowitz Optimization Enigma: Is 'Optimized' 'Optimal'?" *Financial Analyst Journal*, 1989: 31-42.

68. Officer, R.R. "The variability of Market factor of New York Stock exchange." *The journal of Buisness*, 1973: Vol. 46, No. 3, pp. 434-453.

69. Pafke, Szilárd & Imre Kondor. *arxiv.* May 20, 2002. http://arxiv.org/abs/cond-mat/0205119.

70. Paiva, O. L. V. Costa and A. C. "Robust portfolio selection using linear-matrix inequalities." *Journal of Economic Dynamics and Control* , 2002: 889-909.

71. PHILIPPE ARTZNER, FREDDY DELBAEN,JEAN-MARC EBER and DAVID HEATH. "COHERENT MEASURES OF RISK." *Mathematical Finance*, 1999: 203-228.

72. Primbs, J. A. ""Portfolio optimization applications of stochastic receding horizon control." *American Control Conference*, 2007.

73. Primbs, JA. "Lqr and receding horizon approaches to multidimensional option hedging under transaction costs." *American Control Conference*, 2010: 6891-6896.

74. Richards, A. G. *Robust model predictive control.* Ph.D. dissertation, Massachusetts Institute of Technology, 2002.

75. Robin Bade, Michael Parkin. *Foundation of Economics.* Addison Wesley Publishing Company, 2003.

76. Rockafellar, R. T. & Uryasev, S. "Conditional Value-at-risk For General Loss Distributions." 2001.

77. Roley, D. K. Pearce and V. V. "Stock prices and economic news." *The Journal of Business*, 1985: 49 – 67.

78. Roley, Douglas K. Pearce and V. Vance. "Stock Prices and Economic News." *The Journal of Business,Vol. 58, No. 1*, 1985: 49-67.

79. Roll, Robert Geske and Richard. "The fiscal and Monetary linkage between stock Returns and Inflation." *Journal of Finance*, 1983: Vol 38.

80. Samuelson, P. A. "Lifetime portfolio selection by dynamic stochastic." *Review of Economics and Statistics*, 1969.: 239 –246.

81. Samuelson, P.A. "General proof that diversification pays." *The Journal of Financial and Quantitative Analysis*, 1967: 1-13.

82. Schewrt, G William. "The Adjustment of Stock Prices to Information about Inflation." *The Journal of Finance*, 1981: 15-29.

83. *SEC.* http://www.sec.gov/answers/treasuries.htm.

84. Shanken, S. P. Kothari and Jay. "Journal of Financial Economics." *Journal of Financial Economics*, 1997: 169-203.

85. Shapiro, S. Basak and A. "Value-at-risk based risk management:." *Review of Financial. Studies*, 2001: 371 – 405.

86. Shephard, A. C. Harvey and N. *Structural time series models.* Amsterdam: Elsevier Science Publishers B V, 1993.

87. Shiller, John Y. Campbell and Robert J. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *The Review of Financial Studies*, 1988: 195-228 .

88. Smith, R. "Robust model predictive control of constrained linear systems." *American Control Conference*, 2004: 245-250.

89. Stefano Cavaglia, Christopher Brightman, Michael Aked. "The Increasing Importance of Industry Factors." *Financial Analysts Journal*, 2000: 41-54.

90. StockCharts. http://stockcharts.com/school/doku.php?id=chart_school:technical_indicators:relative_strength_index_rsi.

91. Sung, J.A Primbs and C. "A stochastic receding horizon control approach to constrained index tracking." *Asia-Pacific Financial Markets* , 2008: 3-15.

92. Treasury, HM. http://www.hm-treasury.gov.uk/data_gdp_backgd.htm.

93. Tsay, R.S. *Analysis of Financial Time Series.* Wiley-Interscience, 2005.

94. U. S. Krokhmal, P. and J. Palmquist. "Portfolio optimization with conditional value-at-risk objective and constraints." *Journal of Risk*, 2001: 42 – 68.

95. V. Hanser, J., J. B. McDonald, and R. D. Nelson. "Some evidence on forecasting time-series with support vector machines." *The Journal of the Operational Research Society*, 2006: 1053-1063.

96. Victor Fang, Chien-Ting Lin and Kunaal M. Parbhoo. "Macroeconomic News, Business Cycles and Australian Financial Markets." *ASIA-PACIFIC FINANCIAL MARKETS*, 2009: Volume 15, Numbers 3-4 ,185-207.

97. Wei-Peng Chen, Huimin Chung,Keng-Yu Ho,Tsui-Ling Hsu. "Portfolio optimization models and mean-variance spanning tests." Prepared for Handbook of Quantitative Finance and Risk Management.

98. Wikipedia. http://en.wikipedia.org/wiki/Monetary_policy.

99. Winkelmann, R. Litterman and K. *Estimating covariance matrices.* Tech Report, Goldman Sachs, 1998.

100. Winker, J. X. Li and P. "Time series simulation with quasi monte carlo methods." *Compuatational Economics*, 2003: 23 – 43.

101. Witsenhausen, H. S. *Minimax control of uncertain systems.* PhD Thesis, Massachusetts institute of technology, 1966.

102. Wolf, O. Ledoit and M. "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection." *Journal of Emperical Finance*, 2003: 603-621.

103. X. Cai, K. Teo, X. Yang, and X. Y. Zhou. "Portfolio optimization under a minimax rule." *Management Science*, 2000: 957-972.

104. X. Deng, Z. Li, and S. Wang. "A minimax portfolio selection strategy with equilibrium." *European Journal of Operational Research*, 2005: 278-292.

105. Yang, K. I. Teo and X. Q. "Portfolio selection problem with minimax type risk function." *Annals of Operations Research*, 2001: 333-349.

106. Yin, X. Y. Zhou and G. "Markowitz's mean-variance optimization with regime switching: A continous time model." *SIAM Journal of Control and Optimization*, 2003: 1466-1482.

107. Young, M. R. "A minimax portfolio selection rule with linear programming solution." *Management Science*, 1998: 673-683.

108. Zhou, D. D. Yao and X. Y. "Tracking a financial benchmark using a few assets." *Journal of Operations Research*, 2006: 232-246.

109. Zvi Bodie, Alex Kane,Alan J Marcus,Pitabas Mohanty. *Investments.* New York: Tata McGraw-Hill Companies, 2006.

APPENDIX A1

INDUSTRIES IN EACH SECTOR

List the industries in each sector

Source : Yahoo Finance

- Basic Materials
    - Agricultural Chemicals
    - Aluminum
    - Chemicals - Major Diversified
    - Copper
    - 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
- Consumer Goods
    - Accessories
    - 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
- House wares & 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 &
- 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
  - Real Estate Development
  - Regional - Mid-Atlantic Banks

    - Regional - Midwest Banks
    - Regional - Northeast Banks
    - Regional - Pacific Banks
    - Regional - Southeast Banks
    - Regional - Southwest Banks
    - REIT - Diversified
    - REIT - Healthcare Facilities
    - REIT - Hotel
    - REIT - Industrial
    - REIT - Office
    - REIT - Residential
    - REIT - Retail
    - 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
  - 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
    - Catalog & Mail Order Houses
    - CATV Systems
    - 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

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

- Utilities

- Diversified Utilities
- Electric Utilities
- Foreign Utilities
- Gas Utilities
- Water Utilities