Aalto University
School of Business

Value Investing with Rule-Based Stock Selection and Data Mining

Management Science

Master's thesis

Ilkka Sistonen

2014

Department of Information and Service Economy
Aalto University
School of Business

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Ilkka Sistonen	
Title of thesis Enhanced Value Investing With Rule-Based Stock Selection and Data Mining	
Degree Master's Degree	
Degree programme Management Technology	
Thesis advisor(s) Markku Kallio	
Year of approval 2014	**Number of pages** 116 **Language** English

Abstract

Paradigm shift is on the way in the financial market and economics theory. Evidence against the previously prevailing assumptions of rationality and market efficiency has become abundant and new models, that are rapidly becoming the main stream, are based on the actual investor behavior that can be empirically observed and provide better fit to the data. The contrast between the old and the new schools of thought serves as a background for this study, and reviewing some contemporary theories and studies of topics such as *enhanced value investing* and *contextual fundamental analysis* justifies its results which under the efficient market hypothesis would be anomalous.

The first objective of this study is to test in the Nordic markets (Finland, Sweden, Norway and Denmark) Joel Greenblatt's investment formula (GF) that he published in his 2005 book: "*The Little Book that Beats the Market*". His method aims to *buy good stocks when they are cheap* and has provided during period 1988-2004 an average 30.8% p.a. return in the U.S. stock market while the S&P 500 index yielded 12.4% annually. Although the average return performance is stellar, the selection rule also accepts some stocks that will be deeply in loss thus making the whole portfolio to suffer periods of underperformance. This undermines the formula's exploitability especially in professional fund management setting. Second objective in this study is to improve GF by developing a model that can filter out the loss producing stocks *ex ante* using information set that is available at the time of investment.

To accomplish the first goal a program that simulates GF investment rule was developed and run with the result of 29.4% p.a net investment return after taxes and trading costs during the research period 2000-2011 while the relevant reference stock index (FTSE Nordic Value Index) returned 7.6% p.a. To develop a model for the second stage, data mining methods were applied in variable and model selection phase. The produced logistic regression model with constant and 8 additional terms (single variables and interactions) was able to predict with 80-90% accuracy the predefined holding period return category for the cases in the research data. Sample data was divided into training, testing and validation partitions to ensure the out-of-sample performance of the models. When this model was applied as a filter in the stock selection phase the annual return for the period increased to 43.8 % p.a.

KEYWORDS stock investing, Joel Greenblatt, Magic Formula, enhanced value investing, data mining

Aalto-yliopisto
Kauppakorkeakoulu

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Maisterintutkinnon tutkielman tiivistelmä

Tekijä Ilkka Sistonen

Työn nimi Parannettu arvosijoitusstrategia säännönmukaista osakevalintaa ja tiedonlouhintaa hyväksikäyttäen

Tutkinto Kauppatieteiden maisteri

Koulutusohjelma Liikkeenjohdon systeemit

Työn ohjaaja(t) Markku Kallio

Hyväksymisvuosi 2014 **Sivumäärä** 116 **Kieli** Englanti

Tiivistelmä

Rahoitus- ja taloustieteissä on parasta aikaa käynnissä murros, jossa vanhoista rationaalisuuteen ja markkinoiden tehokkuuteen perustuvista teorioista ollaan siirtymässä tutkimaan todelliseen sijoittajien ja kuluttajien käyttäytymiseen perustuvia malleja. Nämä mallit sopivat paljon edeltäjiään paremmin yhteen havaitun datan kanssa. Vanhan ja uuden koulukunnan näkemysten väliset eroavuudet tarjoavat sopivan taustan tälle tutkimukselle, jonka tulokset olisivat vaivoin sovitettavissa yhteen tehokkaiden markkinoiden hypoteesin kanssa. Viimeaikaiset tutkimukset aiheista kuten parannettu arvosijoittaminen ja kontekstuaalinen perusteanalyysi, joihin tekstissä viitataan, sen sijaan ovat tälle tutkimukselle relevanttia viitekehystä.

Ensimmäisenä päämääränä tällä tutkimuksella oli testata Pohjoismaisilla markkinoilla käsittäen Suomen, Ruotsin, Norjan ja Tanskan pörssit, Joel Greenblattin vuonna 2005 julkaisemassa sijoituskirjassaan ”*The Little Book that Beats the Market*” esittelemä sijoitusmenetelmä (GF). Menetelmän tarkoituksena on ostaa hyviä osakkeita silloin kuin ne ovat edullisia, ja kirjassa esitellyssä tutkimuksessa vuosien 1988 – 2004 välisenä aikana GM:n keskimääräinen vuosituotto Yhdysvaltojen markkinoilla oli 30.8% kun samaan aikaan S&P500-Indeksin tuotto jäi 12.4%:n p.a.. Vaikka sijoitusmenetelmän keskimääräinen tuotto on loistava, osa valintasäännön mukaisesti salkkuun ostetuista osakkeista on pitoaikanansa raskaasti tappioillisia. Tällä on se vaikutus, että koko mallinmukainen sijoitussalkku jää ajoittain jälkeen verrokki-indeksistään heikentäen sen käytettävyyttä etenkin ammattimaisen salkunhoidon apuvälineenä. Tästä syystä toisena tutkimuksen tavoitteena oli kehittää malli, jolla voitaisiin suodataa sijoitussalkusta pois huonosti tuottavat osakkeet etukäteen käyttäen sijoitushetkellä saatavissa olevaa informaatiota.

Ensimmäisen tavoitteen saavuttamiseksi kehitettiin ohjelma, joka simuloi GF:n mukaisen sijitustoiminnan vuosina 2000-2011. Tuloksena oli vuosittainen 29.4% nettotuotto sijoitukselle verojen ja kulujen jälkeen. Samana aikajaksona verrokki-indeksi (FTSE Pohjoismainen Arvo-osake Indeksi) tuotti 7.6% p.a. Toisen vaiheen päämäränä olleen pitoajan tuoton ennustefunktion kehittämiseksi käytettiin tiedonlouhinnan menetelmiä, joiden tuloksena saatu logistinen regressiomalli pystyi 80-90% tarkkuudella ennustamaan jokaiselle otannan osakkeelle oikean kategorian pitoajan tuoton suhteen kolmijakoisella luokittelulla. Kun tätä mallia käytettiin suodattamaan osakevalintoja parani GF:n vuosituotto 43.8% p.a.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

LIST OF ABBREVIATIONS

APT	Arbitrage Pricing Theory
BFMA	Behavioral Finance Macro
BFMI	Behavioral Finance Micro
C/P	Cash Flow to Price
CAPM	Capital Asset Pricing Model
CCA	Contingent Claim Analysis
CCAPM	Consumption CAPM
CDO	Collaterized Debt Obligation
CDS	Credit Default Swap
CPT	Cumulative Prospect Theory
CVaR	Conditional Value-at-Risk
DM	Data Mining
EBIT	Earnings Before Interest and Taxes
EM	Expectation Maximization
EMH	Efficient Market Hypothesis
ES	Expected Shortfall
ETL	Expected Tail Loss
EU	Expected Utility
EUT	Expected Utility Theory
EVT	Extreme Value Theory
EVVaR	Extreme Value VaR
GBM	Generalized Behavioral Model
GF	Greenblatt's Investment Formula

HPR	Holding Period Return
ICAPM	Intertemporal CAPM
IMRS	Intertemporal marginal rate of substitution
LR	Likelihood Ratio
MAR	Missing at Random
MCAR	Missing Completely at Random
MFT	Modern Financial Theory
MPT	Modern Portfolio Theory
NMAR	Not Missing at Random
OPT	Option Pricing Theory
P/B	Price to Book
P/E	Price to Earnings
PCA	Principal Component Analysis
PT	Prospect Theory
RDEU	Rank-dependent Expected Utility
ROA	Return on Assets
ROE	Return on Equity
SDF	Stochastic Discount Factor
VaR	Value-at-Risk
VIF	Variance Inflation Factor

1. INTRODUCTION

Where do the asset prices really come from? What are the factors and forces that drive them? Financial academics and practitioners has centuries sought to answer those questions and find a way to gain extraordinary profits by somehow predicting future asset prices or exploiting some style based investment strategy. Others say that it is not plausible and claim that markets are efficient and investors rational to a degree that eliminates the possibility of above market returns. This hypothesis has many shortcomings and growing number of researchers are producing results that prove the opposite. In this thesis I first review some theories and studies that indeed show that markets are not efficient in the traditional sense, and the simplistic models that are in past used to determine asset prices are not adequate for practical work I show that even a famous and published investment rule can produce returns that are well above those that proponents of index investing say is possible. I also review and apply some recent results in data mining (DM) related prediction variable search techniques in context of building economic and financial models.

1.1 Background

Finance theory has in recent decades been divided to two distinct schools of thought over whether capital markets are rational or does human cognitive psychology have influence in asset pricing. The former are called traditionalist or neoclassicists and the latter behaviorists. Traditionalists base their view in assumption of rational capital markets where investors are risk-averse von Neumann-Morgenstern expected utility maximizers with the ability to make unbiased forecasts of the future, and market prices continually reveal the true values of assets.

Under the Efficient Market Hypothesis (EMH) proposed by Fama (1970) information efficiency of markets can be categorized with three level system (weak, semi-strong and strong form), and that they meet at least the semi-strong criteria of effectiveness, meaning that all published information without delay incorporates itself in observed market prices. There is no possibility to consistently make extraordinary profits with technical (weak form) or fundamental analysis, without insider information (strong form). The assumption of effective markets is the fundamental theory underpinning Modern Portfolio Theory (MPT) (Markowitz 1952) and separation theorem (Tobin 1958), that provide for risk-averse investors with mean-variance preferences the optimal solution for asset allocation. MPT assumes that investors have homogenic presumptions of assets risk and return characteristics. For the same ground are built the asset pricing theories The Capital Asset Pricing Model (CAPM) (Sharpe 1964; Lintner 1965; Mossin 1966) ,with its many successors that I discuss more later, and the Arbitrage Pricing Theory (APT) (Ross 1976) with factor models, although the last mentioned are less stringent with assumptions of investor utility.

Behaviorists represent an alternative school of thought in investment research. They add human psychology into equation and claim that investors often are irrational, and make systematically biased decisions based on feelings and heuristics. They use crude rules of thumb for making judgments about probabilities, statistics and future outcomes rather than thorough analysis. These deviations from rationality gives rise to a range of anomalous market and stock return characteristics that the traditional theory and models are not able to explain. The behavioral approach to investing also makes use of the anomalies and claims that many of them are persistent and hence exploitable for making excess profits. For example of biases are **(1)** overconfidence to own skills (Odean 1998), **(2)** under- and overreaction to new information which gives rise to long term reversals (negative long term autocorrelation) and short term trends (positive short term autocorrelation) in stock prices which thus become partly predictable violating the EMH (e.g. De Bondt and Thaler 1987; Jegadeesh and Titman 1993), **(3)** loss aversion which means being more sensitive to losses than wins (Kahneman and Tversky 1979), **(4)** extensive extrapolation of past performance to future (e.g. De Bondt 1993), **(5)** herd behavior i.e. moving with others in markets instead of own beliefs and analysis (e.g. Avery and Zemsky 1998; Lee 1998) and many others that all are today well documented phenomena.

Even fund managers are susceptible to herding bias that leads them to sell recent losers and buy winners in suboptimal way (Scharfstein and Stein 1990; Lakonishok et. al.1992). Welch (1998, 2001) also provides evidence that more than 200 finance professors predicted high stock returns after long bull market in late 1990s and low returns after crash in 2001. This human behavior leads to mispricing of assets that can be persistent due to several limits to arbitrage that prevent rational traders from exploiting the inefficiencies (Shleifer and Vishny 1997). Behavioral school presents a descriptive approach to investment research meaning that they study what investors *really do* versus what they are supposed optimally do according to some hypothetical models. The traditionalist method is prescriptive and it first builds models that are based on a series of assumptions then studies the optimal investor behavior under those conditions.

Although behavioral finance is relatively new as a discipline, it has its roots in economics since Adam Smith's *The Theory of Moral Sentiments* and more recently in e.g. Keynes' work where he claims that it is the human psychology that drives economy into booms and busts. One of the first books that straight applied psychology to stock market was Selden's 1912 *Psychology of the Stock Market*, which was preceded by study of group market's group behavior Mackay's *Delusions and the Madness of Crowds* published in 1841. The inability of the traditional neoclassical finance theory to produce explanations to some well researched stock market anomalies e.g. excess volatility, long run reversals in share prices, momentum and value effects, led to think that investor sentiment may affect equity returns in a way that sophisticated investors may be able to systematically benefit from other people's cognitive and emotional shortcomings. Not to mention the crashes and bubbles that far more often than would be statistically plausible, dislocate asset prices. The recent rise of behavioral finance started in the 1980s with Prospect Theory (PT) of Daniel Kahneman and Amos Tversky (1979) that provided psychologically more accurate description of preferences compared to expected utility theory and was able to explain away many of the problematic issues of rational market school. PT earned Kahneman 2002 Nobel Prize in Economics. His partner Tversky had passed away some years earlier.

1.2 Research Problem and Objectives

Joel Greenblatt, a famous value investor and academic, published 2005 a book called: "*The little book that beats the market"*. There he describes a simple rule to pick stocks based on quarterly published financial statement information. His method aims to buy good stocks when they are cheap and has provided during period 1988-2004 an average 30.8% p.a. return in U.S. stock market while S&P 500 index yielded 12.4% annually. Greenblatt argues that the success of his model is connected to some systematic behavioral biases in markets, and it will continue to work on average. But there is a problem that many of the stocks selected by GF perform very badly while some others are extreme winners that multiply in value. This may lead to periods of inferior performance that easily depletes investors' faith in the system. Greenblatt is founder of Gotham Capital hedge fund that returned over 40% p.a for twenty years in a row and a Columbian Business School professor so his claim that the Greenblatt's Formula (GF) will work also in the future is interesting enough to test in Nordic markets (Finland, Sweden, Denmark and Norway). This study does that as its first objective and in addition improves the formula by developing a model that can statistically classify the stocks that are selected by GF to winners and losers *ex ante* using information set that is available at the time of investment to control for the above mentioned wide dispersion in the holding period return percentage (HPR-%) of the invested stocks.

The use of firm's characteristics, financial statement data and macroeconomic variables to predict its equity performance has long traditions in finance research. Notably the famous Fama and French 1992, 1993 and 1996 papers, where they pronounced CAPM beta dead and suggested several other factors as better predictors to cross-section of company returns. These papers were followed by a surge in multi-factor return modeling research. This study owes to important paper "*Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers* by Joseph Piotroski (2000), where he succeeds in finding financial statement based variables that can among low P/B stocks aid the selection of winners, improving mean annual return by 7.5 percentage points.

Among others with similar results on *enhanced value strategies* are e.g. Bartov and Kim (2004), who argue successfully that if value phenomenon is due to mispricing, it should be strengthened with the use of some additional criteria; Elze (2010) uses multidimensional value strategies that include also capital return variables which he finds to be most useful in enhancing the simple value strategy. A result that is parallel to Greenblatt's theory where also a modified capital return is used in accordance to value indicator. This research builds on that ground and first attempts to find mispriced securities with the GF method and then, by applying DM techniques, to constitute a filter that can separate future winner stocks from losers in this context. In effect, the search for complementary variables is more systematic than in those above mentioned papers, and it is taken beyond simple intuition by using DM. Earlier DM was usually not openly applied to model and variable selection due to problems in statistical inference but as I will later show these problems are not so restrictive anymore as the theory have advanced. DM methods are fast becoming an established part of applied finance and econometric research, like they already are in many other fields of science, in discovering new information from data and formulating hypotheses based on it. This study contributes to existing literature by showing how DM can be applied in practical financial engineering problem.

To put it succinctly, the purpose of this study is to **(1)** test if GF works also in the Nordic markets and provide some explanations to its extraordinary returns by connecting it with behavioral finance research (later 1st stage analysis) **(2)** improve GF's performance and exploitability by finding a pattern from investment time data set and a model, applying DM methods, that can separate winner stocks from losers in the context of stocks that are selected by this investment rule (2nd stage analysis). I also justify the methods used by reviewing some other papers with related approaches like *contextual fundamental analysis* and *automated knowledge discovery and inference in econometrics*.

1.3 Structure and Limitations of the Study

One possible limitation has its origin on the relative scarceness of available detailed financial statement data for Nordic companies hence the period studied is shorter than in Greenblatt's book with U.S. stocks. There were very few data points available before 1Q2000 and they remained sparse also some following quarters. So as the coverage would ideally have included all the exchange traded non-financial companies in every quarter, it is now slightly more limited than that. Still there is no reason to suspect any systematic bias that would interfere with the conclusions of the study. Another possible limitation is that the test is only conducted for the long side i.e. with stocks bought. It would be interesting, also from commercial applications' view, to know how the very low ranking stocks position in future returns, and do they, as a group, form a well behaving (controllable) or homogenous modeling context for classification like they do in the high ranking region. This is left for future research since, at the timeline this data represents; it would have been unrealistic to assume short selling to be allowed in the Nordic exchanges.

As for the structure of this study, in the next chapter I will provide a non-comprehensive and non-mathematical review of relevant theories and literature putting some weight in the disagreement between the neoclassical and the behavioral finance schools of thought as it is considered important for establishing the background for the study. Then the data issues are discussed more thoroughly in Chapter 3 accompanied by detailed explanation of GF and its application. Also in Methods and Data is primer to DM methods used in this study with descriptive statistics of the stocks involved in the 2nd stage analysis. Chapter 4 is for presenting the results of both the GF test and the 2nd stage analysis with the classification model and its performance statistics.

2. LITERATURE REVIEW

In this chapter I review the development from modern portfolio theory to behavioral finance and discuss some of the differences between these two disciplines in finance research tradition. The chapter is arranged so that section 2.1 narrates the neoclassicists' fundamental tenets emphasizing their inability to stand the test of reality which has exposed a series of anomalies in financial asset returns; 2.2 introduces behavioral finance as the main challenger for traditional finance and presents some behavioral explanations for the anomalies; 2.3 surveys research literature of investment strategies that aim to exploit the market inefficiencies and hence constitute the relevant context for this study; 2.4 then extends the debate to some current topics in post CAPM asset pricing and risk management research.

2.1 Modern Financial Theory

2.1.1 History

One of the earliest neoclassicists and a pioneer of mathematical economics, Irving Fisher, in his book *"The Nature of Capital and Income"* (1906) laid ground for the quantitative finance of the latter part of the century by outlining a course of rational and scientific behavior for stock market participants. In Paris, a French mathematician Louis Bachelier studied the price changes on the stock exchange and realized that the uncertainty and randomness of price changes could be handled with the Gaussian distribution that by the early 20th century was already gaining interest and applications in physics in studying infinite amount of *independent* causes. He was aware of the limitations of its uses in modeling human behavior because, in a group, human actions are no longer independent of each other but rather are prone to heard behavior. In his doctoral thesis which was decades ahead of its time, *"The Theory of Speculation"* (1900), he discovered the mathematics of Brownian motion and stated that past prices cannot predict future, and price changes are best described as random walk. Almost fifty years later Harry Markowitz applied linear programming, that was widely used in operations research, to stock investor's optimal portfolio problem in his article *"Portfolio Selection"* (1952) which built on Fisher's and Bachelier's work.

Cost that he minimized was risk, proxied by variance of returns, and reward by the mean of an asset. Several assumptions, which none actually is realistic, needed to be made for the model to be applicable. Following list is not exhaustive but includes some of the more important ones:

1. **Investors are only interested in mean and variance of return.** In reality investors have utility functions that may be sensitive to higher moments e.g. kurtosis and skewness.

2. **Asset returns are normally distributed.** Large swings (3-6 std.) are far more common that the model would predict.

3. **Investors are only interested in money.** It is likely that some other factors, even qualitative, enter the utility function.

4. **Investors are rational and risk-averse**. Investors are, on average, assumed to process information in a cognitively unbiased, Bayesian fashion. Irrational investors' trades are unlikely to be correlated so they would cancel out. Mounting evidence from the behavioral finance proves that investors are just the opposite.

5. **All investors have access to the same information at the same time. They have homogeneous expectations of the risk–return characteristics of assets and know their probability distribution, which stays constant, in advance.** In reality information is asymmetric in many ways. It is not realistic to assume that small investor and corporate insiders and investment bank analysts are equally informed. People do have different opinions about outlook of stocks, and their expected parameters change in time as new information arrives. The assumption that everybody knows the underlying return-generating process, its parameters and functional form and agrees on it for each stock is not realistic.

6. **There are no limits to arbitrage.** In theory when price drifts away from intrinsic value the informed traders should step in and arbitrage away the mispricing by trading against the uninformed traders (noise traders). In practice there are several limitations to this process (Shleifer and Vishny 1995). *"The Markets can remain irrational longer than you can stay solvent"*, a quote from John Maynard Keynes which truthfulness is evidenced e.g. in the collapse of LTCM 1998 in the turbulence of the Russian default.

7. **Correlations between assets are fixed and constant forever.** Correlations depend on systemic relationships between the underlying assets, and change when these relationships change.

8. **Any investor can lend and borrow an unlimited amount at the risk-free rate of interest**. This is another false assumption based corner stone of modern finance. Modigliani–Miller theorem states that given that market prices follow random walk, in the absence of taxes, bankruptcy costs, agency costs, asymmetric information, and in an efficient market, the capital structure does not affect the firm's valuation. This idea was soon combined with tax-deductibility of interest expenses which led in theory to arbitrarily high optimal leverage. The kind of thinking that derives from the combination of these two ideas has probably been one reason that led the financial markets to the use of excessive leverage and thus is an important factor behind its increasing instability. In reality the access to capital dries out immediately when there is a sign of trouble, and it is impossible to renew limits, loans and any sort of external financing. Usually this happens when the liquidity in asset market also vanishes and unwinding of positions can (if at all) be executed only with deep discounts. So the costs of financial distress can be very severe but the theory does not include them.

9. **All investors are price takers, i.e., their actions do not influence prices.** This is often not true with sufficiently large institutions in some exotic instrument market. Especially if there is turbulence. This, like many other assumptions, brake down exactly when they would be most needed to hold.

Markowitz realized that variance was not necessarily a correct measure for asset's riskiness, and he proposed that semi-variance, downside risk, would be more realistic choice. Also, he knew that not everybody was just maximizing the mean of returns, but also the higher moments mattered (Markowitz 1959). The model thus got its final shape much because the need for mathematical elegancy and ease of calculation considering the limited computer resources of that time. Markowitz's model was a prescription for how investors should behave optimally but Jack Treynor, his friend John Lintner and Markowitz's student William Sharpe took it as a description of actual investor behavior and created an equilibrium asset pricing model CAPM which, compared to Markowitz's Modern Portfolio Theory (MPT), had great advantage in estimation since the assets' pairwise covariances were no longer needed. Then Eugene Fama published his thesis about the random walk hypothesis (1965) where he stated that successive price returns are independent over time. Paul Samuelson proved efficient market hypothesis which Fama (1970) backed with evidence and extended to include the definitions of the three forms of market efficiency.

After that the term *efficient market* has been most often used to describe markets where no investor can outperform his rivals by consistently generating abnormal risk-adjusted returns. The trading mechanism would immediately correct all deviations from the assets true value. This kind of neoclassicist thinking led to a surge for the index fund industry starting from the late 1970s and continuing until today. As it became the mainstream view among academics and practitioners, it also had a significant effect on economic policy as a whole. Supporters of this efficient and rational market school of thought presented strong arguments against central bank and government regulation of the financial markets and claimed that, if left alone, the rational markets would always stabilize to equilibrium level of asset prices. This part of the theory actually origins from Adam Smith, the 18^{th} century economist, who is often said to be the founder of modern economics. He described the mechanism of free markets as an invisible hand and his recommendation for regulators was *laissez-faire* – leave it alone.

2.1.2 Anomalies

Anomaly is an unexpected price behavior in equity markets i.e. it is not captured by CAPM. If it is persistent in the sense that it does not seem to vanish after investors start exploiting it, either it is market inefficiency which means profit opportunity or the underlying asset pricing model is inadequate. When researchers find a new anomaly that seems to be persistent, the proponents of the EMH usually argue that the related extraordinary returns are a reward for an exposure to some additional fundamental risk factor (other than market beta) to the asset pricing model and denying the possibility of market inefficiency due to e.g. human psychology or limits to arbitrage. The anomalies based literature distinguishes these four types of cases (Diamond and Zacks 2011, p. 5-9):

1. Is the anomaly real or is it an artifact of mismeasured risk or resulting from erroneous statistical inference? Test of market efficiency is always jointly a test of assumed underlying asset pricing model. Therefore a failure of the joint hypothesis could be due to a misspecification of the expected return model, rather than to failure of market efficiency (Fama 1970).

2. Is the anomaly a result of uncertainty about the underlying return-generating process? This would be a violation of assumption 6 of efficient markets in previous subsection. This is most likely to happen with new, small or otherwise less analyzed stocks. Empirically most anomalies are generated by the returns of this category firms (e.g. Fama and French 1993; Merton 1987; Brav and Heaton 2002).

3. Is the anomaly result of investors' psychological biases? This would violate assumptions 1, 3 and 5 listed above. Behavioral finance is a class of theories that relax the rationality assumption and explain many of the observed anomalies with behavioral or cognitive biases of investors.

4. A fourth subset of the literature explores whether limits to arbitrage can explain the persistence of mispricing. This is an important argument of behaviorists and a violation of the 6th assumption of effective markets stated earlier.

While general belief in the rational market theory holds on at least until the 1990s, the amount of anomalies that could not be explained by rational markets or CAPM started to grow already in 1970s and 1980s. Already in 1977 Basu proved that low P/E stocks generate higher average returns that high P/E stocks. This anomaly was named as value effect and was soon confirmed by dozens of other studies where returns are predicted by ratios of market value to accounting measures such as earnings or the book value of equity (e.g. Basu 1983; Rosenberg et al. 1985; Fama and French 1992). Banz (1981) reported the size effect that low market cap stocks have higher average excess returns than can be explained by CAPM. There has been a lot of arguing whether these effects are really symptoms of mispricing or are they just rewards for some underlying risk factors. Logical claim from EMH supporters is that this value premium, which existence cannot be denied, is just demonstration of rational risk pricing. This latter view has been advocated by e.g. by Fama and French in series of published papers in years 1992- 2006, and who as supporters of the EMH, claim that the value premium is compensation for higher fundamental risk factors like financial distress. However, Dichev (1998) and Campbell et al. (2008) show that distressed firms tend to have lower than average future stock returns, hence making the risk explanation unlikely. And in fact, if the distressed companies are separated from the low P/B stocks, the value premium for low P/B is even higher. Griffin and Lemmon (2002) find that the value effect, measured by return difference between high and low P/B stocks, is twice as high among distressed stocks than other stocks, and distressed stocks have negative pricing error (alpha). And further many other researches has come to conclusions that value stocks are not overall riskier than growth stocks but, in fact, exactly the opposite (e.g. Chan and Lakonishok 2004; Magnuson 2011). Kwag and Lee (2006) studied the growth and value stock returns over business cycle and concluded that risk corrected value stock returns were always better than growth stock and this was especially true during economic downturns. On the other hand, several behavioral explanations for value premium and other anomalies have merged.

(1) Investors are prone to extrapolate past too far into future i.e. buying past gainers and selling losers (Lakonishok 1994) and end up consistently overpaying for glamour stocks (successful growth companies) that subsequently fail the expectations (e.g. Kahneman and Riepe1998; Gilovich, Griffin and Kahneman 2002); **(2)** investors have tendency to overreact in short term and underreact to information in long term; **(3)** there exists agency costs attached to professional money managers based on that they are found to be biased to investing in glamour stocks for reasons concerning their professional reputation (Lakonishok et al. 2004). The findings of DeBondt and Thaler (1985) show that stocks with low returns over the past three to five years outperform in the future and *vice versa*, hence stocks make long swings around their true value overshooting up and down for reasons they relate to investor behavior. Jegadeesh and Titman (1993) documented a momentum effect in which stocks with high returns over the past three to 12 months tend to outperform in the future (3-12 months). Long-reversal and short term momentum patterns strongly contradict the unpredictability of stock prices, and thus market efficiency, since no risk-based story for their existence has yet emerged.

2.1.3 End of Modern Financial Theory

Soon after its emergence CAPM became the predominant theory of asset pricing. It was fairly simple to use and teach, and it was able to produce strong testable results and predictions for assets' risk-return relationships which made it popular among academics and finance industry. Despite of the growing criticism, it maintained that position at least until the mid-1990s. Still it is usually the only asset pricing model taught in undergraduate and MBA finance courses, and it is commonly used in applications like estimating the cost of capital for firms and evaluating performance of managed portfolios.

Early tests of CAPM already rejected the Sharpe-Lintner version of the model finding that alpha-parameter, the intercept, was higher than risk-free interest rate and the relation between the beta and average return was too flat compared to model prediction (Black, Jensen and Scholes 1972; Fama and MacBeth 1973). These results went largely unnoticed and did not have any effect of the popularity of the model. First widely noticed critique was Roll (1977) where he claimed that the true market portfolio was unobservable and thus CAPM was in practice useless.

Fatal blow came from the father of the EMH, Eugene Fama, who early on was generally supporting the model. He stated that the relation between the beta and asset returns was very weak at best and included other factors like value and size effects into the model producing empirically much better fit to cross-section of asset returns. Literally, he declared that beta is dead (Fama and French 1992). He also criticized some of the model assumptions as being unrealistic e.g. the investor's interest only to mean and variance of portfolio and one period investment horizon. His finding was that it is likely that investors also are interested in their portfolio's covariance with their human capital and future investment prospects and thus the CAPM beta lacks the true dimensionality of investment risk. His conclusion is that the use of CAPM, in other than teaching the principals of portfolio theory, is not recommended. Its continued use can best be derived from its usefulness to finance industry (Fama and French 2004).

Still the modern quantitative finance, as practice of reliance on theoretically elegant mathematical models based on EMH and list of other unrealistic assumptions, proved to be very resilient and continued unchallenged despite the fact that most of the fine-tuned mathematical models work when markets are calm and brake down exactly when the turmoil starts sending most asset correlations approach to unity. Here the convenient assumption of Gaussian return process is destroyed along with illusive safeguard functions of Markowitz style portfolio diversification which again is based on historical parameter estimates. One plausible explanation is that as long as everyone else is using these respectably scientific looking and generally accepted models, a risk manager is professionally protected when the accident happens. Keynes ones defined a sound banker as *"not one who foresees danger and avoids it, but one who when he is ruined, is ruined in a conventional and orthodox way along with his fellows, so that no one can really blame him."* (Keynes 1931, p.176)

Since Fisher's days it was believed that the stock prices in rational markets reflect the expectations of present values of future dividends. First to cast a serious doubt over this hypothesis was Robert Schiller (1981) who claimed that in reality market volatility far exceeds the economic explanation and thus EMH. An independent study of Leroy and Porter (1981) came to same conclusions and added weight to irrationality-inefficiency theory advocates' arguments.

A year earlier published paper by Grossman and Stiglitz (1980) titled *"On the Impossibility of Informally Efficient Markets,"* raised a question of why anybody would bother to analyze stocks if there were no compensation to be gained. Anyway the efficient market school flourished and new computerized and mathematical trading models were developed on the assumption that the EMH holds. Mathematicians, physicists and finance professors were recruited to hedge funds and investment banks in an effort to, paradoxically against their tenets, beat the market. One of those new quantitative models was portfolio insurance that soon, in the early 80s, became very popular among institutional investors e.g. pension funds since the hedging could be done using solely index derivatives. Then on October 19th 1987, the model broke when large enough quantity of selling pressure hit the market simultaneously resulting in a worst single day crash in U.S. stock market history. Both the Dow Jones Index and S&P500 lost in excess of 20% of their value, and efficient market skepticism surely begun to gain foothold among market participants.

There were a number of problems with the portfolio insurance: it depended on a string of assumptions that did not hold in reality (e.g. perfect market liquidity, the ability to trade quickly at low cost). It was also intuitively misguided strategy, in the sense that it called for investors to sell when the price had fallen, and buy back when the price had risen, whereas usually profits are made by the opposite strategy. But perhaps most of all, portfolio insurance requires that traders using it be price takers unable to affect the market price. Once too many people use it, they affect the price itself, and the strategy no longer works, even in theory. After the crash Mark Rubinstein, a famous finance professor in the field of derivatives, developer of the binomial option pricing model, and indeed, developer of the portfolio insurance financial product, calculated that probability for such a drop, in a world of normally distributed price changes, was around 10^{-160}. That is, it was something investors could expect to happen once every couple billion billion years. The universe has only existed an estimated 12 billion years; the New York Stock Exchange was, as of October 1987, 170 years old (Jackwerth and Rubinstein 1996). Robert Shiller, an influential economics professor in Yale and bestselling book author, among being 2013 Nobel laureate in economics, said after the Black Monday, *"The efficient market hypothesis is the most remarkable error in the history of economic theory. This is just another nail in its coffin"* (Fox 2011, p.232).

All the statistical quantitative finance models of The Modern Financial Theory (MFT) are based on the normality or at least log-normality (continuously compounded returns are normally distributed) of gross asset returns. Empirically the daily stock market return data does not well fit to this hypothesis but is more leptokurtic (peaked-in-mean and heavy-tailed) than log-normal distribution. Essentially this means that the multiple sigma swings (tail-events), like the Black Monday above, are far more frequent than in normal distribution. Derivative traders learned their lesson, and phenomenon called volatility smile, where far out of money options trade with higher implicit volatility, emerged in options markets after the event. There is no fundamental based theory that would presuppose asset returns to obey a certain distribution so the question is purely statistical and empirical. Empirical research has found at least the following stylized facts in market returns: **clustering of volatility –** large movements are followed by large movements, **autoregressive behavior –** price changes are dependent on the past changes, **skewness –** the potential scale of the up and down price changes is asymmetrical, **fat tails –** the probability for very large price changes to both directions is larger than predicted by the normal distribution. An acceptable risk measure must be able to capture these stylized facts (Stoyanov et al. 2011). Early advocates of using so called stable Paretian family distributions (also called Levy distributions after their inventor mathematician Paul Levy) to examine financial data were mathematician Benoit Mandelbrot (1963), famous for his applications of fractals and chaos in finance, and Eugene Fama (1965) who argue that these distributions (normal distribution is a special case) generally fit much better to the data and can accommodate both fat tails and asymmetry. Fama and Roll (1968), So (1987) and Mantegna and Stanley (1995) and others have since found support for stable Paretian data generating processes in a wide variety of financial time series. **Figure 1** has examples of different parameterizations of Levy distributions. The distribution is described by four parameters: α– the characteristics exponent, β– a skewness parameter, c– scale parameter and μ- a location parameter. Mandelbrot's discovery was initially well received, and he was offered a professorship at the University of Chicago Graduate School of Business but the offer was withdrawn, and his work was dismissed, ones implications of his theory to MFT were understood. Stable Paretian distributions are much more demanding to estimate, and there is a big problem of infinite variance when $\alpha < 2$.This simply means that when random observations are repeatedly drawn from stable Paretian distribution and added to the existing sample, the sample estimate of standard deviation that is recalculated after each addition, is then likely to grow without a limit. In fact, the only distribution with finite variance within this family of distributions is the Gaussian (α=2).

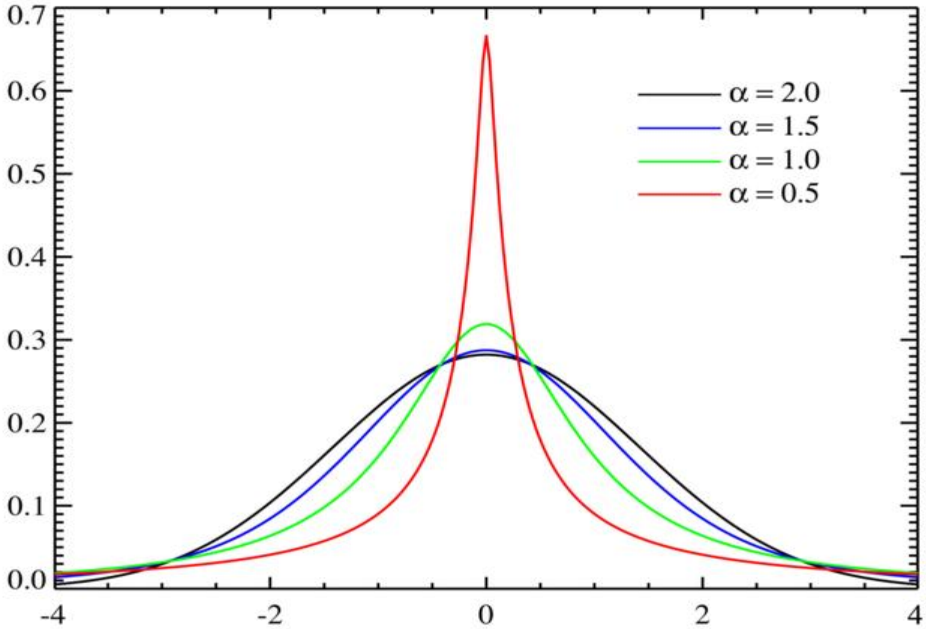


Figure 1: Levy Distribution (Wikipedia)

Firms' investment and financing decisions will affect the systematic risk, expected return, and standard deviation of returns. Macroeconomic fluctuations will also affect to the level of interest rates and market risk premiums. Therefore, it is not surprising that the risk–return characteristics of single stocks and portfolios are dynamic. So the above mentioned non-Gaussian characteristics of empirical returns could easily be caused by non-stationarity of the distribution's two first moments (Francis and Kim 2013, p.201). The issue is still debated but never the less either solution does not work with EMH, MPT, CAPM or Black & Scholes -formula which all require the variance to be finite and consecutive price changes independent and smooth. This is not true if the stock returns follow the stable Paretian process where large jumps both up and down occur frequently producing discontinuities.

This jump behavior greatly increases the riskiness compared with the Gaussian return generating process. Risk-averse investors thus participate in stock markets more reluctantly and demand higher risk premium than standard models predict (Fama 1963). According to the EMH, this would imply that also the firms' intrinsic values often shift violently in very short periods in order to justify the market effectiveness despite the high price volatility (Fama 1965). The added realism was just not enough to compensate the harmful impacts to finance industry, and the consensus remained that the old models were to hold.

The crash of 1987 was the first alarming demonstration of inherent instability of the mathematical risk-management models of MFT. Next incident came 1998 when LTCM collapsed in spectacular way having not one but two Nobel laureates on board. Myron Scholes and Robert Merton both received Nobel prices in Economics for their contributions in derivatives pricing theory. This time no one could blame incompetence for the failure – the flaw would have to be in the underlying assumptions LTCM based their trading strategy. They pair traded small differences between related instruments that were supposed converge in efficient markets, and levered their position in scale 50-1. This was not supposed to be a problem since, as William Sharpe characterized LTCM for the Wall Street Journal, *"It was probably the best academic finance department in the world"* (Mandelbrot and Hudson 2010, p.106). Then something out of the ordinary happened (Russian default), and their strategy failed as contrary to their expectation, the spreads started to widen. There was no way to unwind their position since other market players knew they were in trouble and financial markets started to play strategic games instead of functioning as an efficient price formation apparatus. LTCM soon lost almost all its capital, and its leverage went from huge to astronomical until they were saved by Fed orchestrated rescue plan. This is a good example of how the deceivingly solid looking quantitative trading strategies of MFT reduce to picking up pennies in front of a steamroller. It means enjoying small gains from taking huge hidden, unknown and uncompensated risks (e.g. correlation risk, liquidity risk) that in reality can wipe you out after every few years. This is a cause of the inherently behavioral nature of the financial markets that is unaccounted in the models.

Some time passed again and along came the dotcom bubble-bust. Alan Greenspan made his famous *"irrational exuberance"* speech in 1996 in an effort to restrain the price rally but as devoted EMH worshipper, did nothing, and also remained in sidelines allowing markets frenzy without restrain for a good three more years until he hit the brakes with a series of interest rate hikes, and the bubble bursted. NASDAQ index rose from 800 in the mid-1990s, to more than 5000 at its peak in March 2000 when it collapsed suddenly, falling below 2000 next year and until to little over 1000 a year after that. It was obvious that the prices being paid for dotcom investments could not be justified on the basis of the prevailing standard principles of valuation. EMH had failed the test again. According to its principles arbitrageurs, the informed traders should have gone against the irrational market crowd (noise traders) and sell short the overpriced companies. As it happened there were only few who could do that and survive; many, like Julian Robertson of Tiger Investments lost billions as stocks kept on climbing during 1999 and threw in the towel. A neat citation from Keynes: *"The markets can stay irrational longer than you can stay solvent,"* describes the problem well.

Afterwards theories of rational bubbles, limits to arbitrage, noise trader risks and market sentiment started to emerge in an effort to explain what had just happened. By now one would expect the world to firmly switch over to post EMH era but instead another thing happened. Loose monetary policy after the stock market crash soon started to inflate a new bubble in housing markets. This time the villain was a new highly mathematical quant innovation, Gaussian copula function. A quantitative analyst David Li had come to conclusion that this type of function was the best tool for modeling the joint distribution of underlying default risks of basket credit products. The correlations he estimated, not from historical data, but from the prices of credit default swaps (CDS) which gave an assuring appearance of forward-looking data. The method he produced in his paper *"On Default Correlation: A Copula Function Approach,"* (2000) soon became *de facto* standard in pricing collaterized debt obligations (CDO) in Wall Street, and also the rating agencies followed suit. The Gaussian copula was particularly convenient because it is completely summarized by a correlation matrix that was based on the assumption of multiple normal distribution with stable parameters (Birge and Linetsky 2007, p, 446), and markets were thus expected to move with predictable patterns without portfolio wide extreme moves.

Now, the banks had what they wanted: a nice number that told them there is no risk. As the model was based on the investors' view of housing markets (through CDSs), and housing prices where in rapid rise, Li's formula made these CDO's to look very safe, and this created more demand for mortgage-backed securities. It was a feed-back loop that drove up prices and size of the aggregate market (Patterson 2010, p.171). It was another example of crowding that results from popular quant methodologies; nearly every CDO manager and trader was using the same formula until volatility hit all the markets simultaneously in early 2007, and they started to implode. By 2008 in excess of 2 trillions of dollars worth mortgage- and other asset-backed subprime loans had been wrapped in AAA rated packets and sold to investors or parked in banks' own balance sheets (Triana 2009, p.113). The Fed's view was that markets are efficient, rational and nothing can go wrong. After the markets melted in 2007-2008, crisis had spread globally to all asset markets, the investment banking sector collapsed and the Li's formula was commonly renamed as *the recipe for disaster*. This time there was no economic recession to blame for the crash, the financial markets did it all by themselves and caused the worst financial crises since The Great Depression of the 1930s with flawed models based on faulty assumptions. In fact the financial markets gave a vicious punch to the real-economy which, in theory, they should serve as an efficient and rationally functioning instrument.

One of the culprits to the crises is no doubt Fed chairman, Alan Greenspan, who let the bubble to inflate by keeping the monetary policy loose and refusing to take action to restrict the finance industry's wide spread use of hazardous risk management practices and dubious credit ratings. In 2004, he stated that nationwide and severe bubble in housing markets was very unlikely, his successor Ben Bernanke agreed and 2005 commented on the issue that housing prices are firmly based on fundamentals. Both of them ignored all the warnings and trusted that there cannot be anything wrong because markets are functioning perfectly when left alone. They were not alone of course; majority of the practitioners, scholars and regulators shared this faith.

One of those that tried to ring the warning bell was Robert Schiller, a Yale economics professor and behaviorist researcher who had predicted the 2000 crash in his book *"Irrational Exuberance"* (2000) that was published earlier the same year. 2005 he, in the 2nd edition of the book, added a warning about the housing markets and wrote that if the prices would continue to rise that way, the inevitable collapse would cause a global economic recession. In 2006, he published in the Wall Street Journal an article where he again warned about the looming housing meltdown and following prolonged economic fault condition. Finally in 2007 he wrote that the CDO markets will implode and result in a panic. In Congressional hearing 2008 Greenspan testified that he was in *"a state of shocked disbelief"*, and that he had wrongly trusted in the EMH and rationality of markets i.e. its capability to price risks correctly, adding, in reference to the recent crises and MFT, that: *"the whole intellectual edifice collapsed in the summer of last year"*. When asked specifically if he thinks that the theory that has been dominant last 40 years and guideline in the Fed's policy is flawed, he agreed. In light of the above discussion, it would seem that despite all the quasi-sophistication manifested in supercomputers and increasingly complex mathematical models, the global finance markets are subject to the same manias, bubbles and busts that were seen in the Dutch tulip craze of the 17th century. Perhaps it was time to stop treating finance as one of the natural sciences with exact laws and applying e.g. physics' models to finance as they really could account for complicated human behavior and institutions' strategic decisions.

2.2 Behavioral Finance

2.2.1 Overview

Many empirical shortcomings of the neoclassical economics led to emergence of new kind of school of thought in finance research by the early 1980s. Behavioral finance has two fundamental elements: limits to arbitrage and human psychology (Barberis and Thaler, 2002). Limits to arbitrage means that there, for several reasons, can exist pervasive arbitrage opportunities in asset markets. This means that arbitrageurs and partly rational investors can coexist in the market as the former are not able to fully profit on the market mispricing.

Behavioral finance differs from traditional finance research as it attempts to study actual investor behavior, often in controlled experiments, versus normative and ideal theories of how investors should behave under some assumptions. It discards the concept of rational *Homo Economicus*, the representative agent, and rather claims that investors are prone to make systematic cognitive errors and resort to heuristics rather than careful analysis in decision making.

Market sentiment is the aggregate mispricing effect of correlated errors among market participants (Shefrin 2007). Keynes (1936) had already brought to attention the fact that psychology has a remarkable role in economics. Well before behavioral economics and finance were born as independent disciplines. He used the word sentiment in a meaning of the investing crowd's unrealistic optimism or pessimism which frequently leads to market booms and busts. In his work, he stressed that, such 'moods' of the investors causes that securities prices often diverge from their intrinsic values. This asset market mispricing, on the other hand, has implications for the real economy through employment, income and money. The main argument of behavioral finance is that market participants are not completely rational, market inefficiencies are pervasive, and based on this, market bubbles and crashes inevitably repeat in the economy (Shefrin, 2000). Two common branches of behavioral research can be distinguished: **(1)** Behavioral Finance Micro (BFMI) studies how the perceived behavior and biases of individual investors differs from the rational agent assumed in the classical economics and finance, **(2)** Behavioral Finance Macro (BFMA) focuses on the market anomalies and their behavioral origins (Pompian 2012, p.39).

Two researches from field of cognitive psychology Kahneman and Tversky (1979) used experimental evidence to argue that investors behave as if their utility function is kinked at a reference point which is close to the current level of wealth. Kahneman and Tversky analyzed the decision making under risk, and they proposed prospect theory (PT) which incorporates human emotions to better describe human decision. According to these authors, two phases are involved in the decision-making process: a representation of the facts in the initial phase (framing phase), and valuation phase where the investor evaluates each prospect and chooses the one with the maximum prospective value.

The prospective value function described by them is concave over gains (risk aversion for gains) and convex over losses (risk-seeking for losses), and it is steepest in the loss domain. Moreover, this S-shaped value function reflects the property of diminishing sensitivity. This property is also presented in the weighting function. Under PT, the weighting function overweights low probabilities and under weights high ones. PT integrated a series of insights due to Markowitz (1952), where he ponders why rationally behaving agents would by lottery tickets, and Allais (1953) that presents the so called Allais paradox and experimentally shows that investor behavior deviates from what would be predicted by EUT with a concave utility function.

PT soon became a rich framework and a corner stone for behavioral asset pricing. According to Shefrin (2008, p.392), four main features distinguish it from the expected utility (EU) approach: **(1)** EUT calculates the value as final wealth; and PT assumes that gains and losses relative to a reference point are in reality more relevant; **(2)** EU assumes that investor risk tolerance is fixed whereas PT states that tolerance for risk is different when people perceive themselves to be in the domain of gains than it is when people perceive themselves to be in the domain of losses, and this leads to kinked preference curve like in **Figure 2**; **(3)** whereas EUT postulates that people weight probabilities correctly, PT implies that people overweight some probabilities and underweight others; **(4)** EU sees people indifferent to the situation where the decision is taken, and PT states that people place different weights on gains and losses and on different ranges of probability i.e. frame their decisions. They also found that individuals are much more distressed by prospective losses than they are happy by equivalent gains.

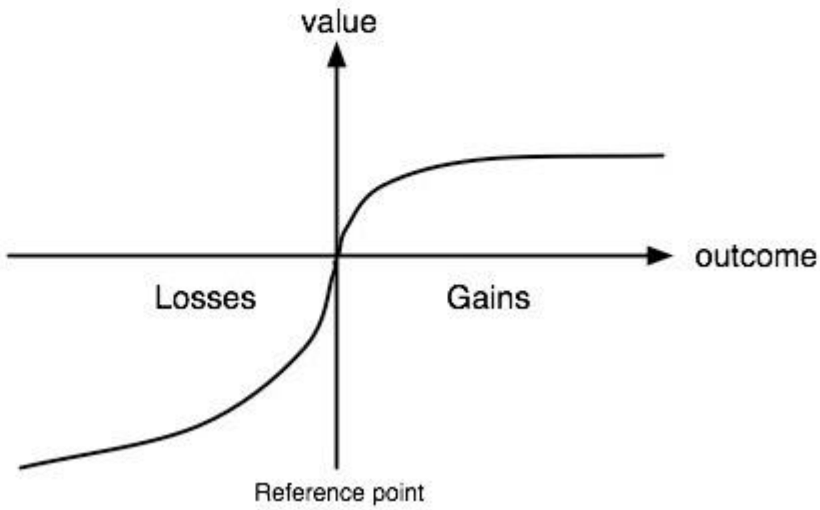


Figure 2: PT Value Function (Wikipedia)

2.2.2 Behavioral biases

Behavioral biases in investment context are irrational financial decisions that are caused by faulty cognitive reasoning or are influenced by emotions. Cognitive errors are symptoms of incorrect reasoning e.g. due to incomplete comprehension of the mathematical process involved in updating probabilities or irrational desire to hold on to ones prior beliefs. Emotional biases on the other hand are more spontaneous and originate from impulse or intuition rather than conscious calculation (Pompian 2012, p.25). Hirshleifer (2001) argues that many or most systematic psychological biases can be viewed as corollaries of heuristic simplification, self-deception, and emotion-based judgments. Systematic here means that they do not cancel out between investors but affect aggregate market prices through limits of arbitrage. Much of earlier work in behavioral finance was to find new anomalies and work out behavioral explanations to them.

Traditionalists responded by trying to prove that the anomalies have already vanished because traders have arbitraged them away, or like Fama did, adding new "risk" factors to asset pricing model. Fama's three-factor, and its extension to four-factor with momentum (Jegadeesh and Titman 1993; Carhart 1997) , do fit cross-section of asset prices better than CAPM, but are rather more examples of applying DM in finance than truly fundamental risk-based models.

Tests of market efficiency are always joint test with model used for expected returns, and Fama as an intellectual father of EMH was more eager to give up CAPM. It was not until 2007 in his article with Kenneth French when he admitted that markets are not as efficient as he had anticipated, and prices could deviate significantly and prolonged periods from fundamentals. There are plethora of identified behavioral biases; in **Table 1** is mentioned a few of the most common with their suggested causes and effects in investment behavior. Adam Smith, who is best known for the concept of the *"invisible hand"* and his famous book "*The Wealth of Nations"* (first published in 1776), wrote a less well-known book "*The Theory of Moral Sentiments"* (1759) in which he stated the psychological principles of individual behavior. The book has lot of insights about human psychology, many of which have a lot in common with modern behavioral economics. For example, Adam Smith commented that *"we suffer more...when we fall from a better to a worse situation, than we ever enjoy when we rise from a worse to a better."* (1759, p.311) That is exactly what PT proposes about loss aversion.

Table 1: Well-Known Behavioral Biases

Bias, type	**Description**	**Decision Effects**
Overconfidence, emotional	Thinking that one's own private information is superior to others.	Has been found in several studies to be a root cause for excessive trading (Benos 1998;Odean 1998a)
Cognitive Dissonance	Belief perseverance bias where investors experience difficulty in absorbing new information that disagrees with past actions.	Tendency to hold on loosing stocks too long (and average down), herding, *"this time it is different"* thinking. In empirical studies of panics and hypes has been found evidence of market reactions towards new information which are consistent with cognitive dissonance. (Kaminsky and Schmukler 1998).
Loss-Aversion, emotional	Negative utility of loss is much stronger than positive from gain.	Hanging on loosing investments and realizing too fast profitable investments (disposition effect).
Representativeness (recency bias), emotional	Investors make decisions based on too small samples and over-weight recent information.	Tendency to extrapolate recent price trend too long disregarding fundamental valuations.
Framing, cognitive	Information processing bias in which investor preferences change as a function of decision context (frame).	Investors may act as if they are risk-averse in some of their choices but risk seeking in other choices.
Anchoring, cognitive	Anchoring is an information processing bias in which the use of psychological heuristics influences the way people estimate probabilities.	Remaining in old target/reference prices like purchase price and disregarding new information.

2.2.3 Limits to arbitrage

The traditional view of markets is that rational speculators (arbitrageurs) stabilize asset prices and counter deviations from fundamentals eliminating irrational traders' (noise traders) destabilizing actions (on average buy when prices are high and sell when they are low) earning positive profit (Friedman 1953). DeLong et al. (1990) find that if arbitrageurs are risk-averse, they are not willing to take big enough positions and thus noise traders can affect prices. They also show that with prevalence of positive-feedback trading (trend following strategies) arbitrageurs can sometimes make rational decision to anticipate noise traders' actions and "jump in the bandwagon" i.e. follow the trend and further destabilize markets with the aim of increased profits at the expense of uninformed feedback noise traders. This view of rational speculation is witnessed in George Soros' (1987) description of his own investment strategy. He claims his success is not based on betting on fundamentals but on future crowd behavior. This model predicts that the short-term correlation of stock returns is positive as feedback traders react to price increase and buy in to the stock. The long-term correlation is, however, negative as the stock price eventually return to its fundamental level. Pattern has been empirically verified in at least Fama and French (1988); Poterba and Summers (1988); Lo and MacKinlay (1988). This model also explains the observed stock market overreaction to news to be driven by the positive-feedback trading. According to Nofsinger and Sias (1999), herding and feedback trading can explain a number of financial phenomena, such as excess volatility, momentum, and reversals in stock prices. They define herding as a *"group of investors trading in the same direction over a period of time"*. Feedback trading that drives prices over the intrinsic stock value is herding that is triggered by the lagged stock returns (momentum signal triggers feedback trading). They find evidence that institutions (informed investors) do engage in positive-feedback trading with significant price effect thus attenuating the ability of arbitrage to correct prices back to fundamentals.

Andrei Shleifer and Robert Vishny (1997) were the first to formally contradict fundamental theory of omnipotent arbitrage in traditional finance's bedrock although DeLong et al. (1990) had already established that it may be rational for arbitrageurs rather to destabilize than stabilize markets. Their study *"The limits of Arbitrage"* shows that arbitrage is usually risky, informed traders (arbitrageurs) are risk-averse and affected by agency issues. This means that the specialized, professional arbitrageurs may be restricted on taking extremely volatile "arbitrage" positions. Although such positions, as buying long a stock and selling short a close overpriced substitute, are potentially very lucrative, and in the long term offer attractive average returns, the volatility that they are exposed to puts them into a risk by their financiers (e.g. investors of the fund or other sort of capital providers) who may require to liquidate the position early in case the price spread widens instead of closes. Shleifer and Vishny claim that the ideal textbook model (and assumption in MFT) where multitude of small, risk neutral arbitrageurs collectively drive prices towards fundamentals does not apply in real markets. The arbitrageurs are more likely to be a fairly small group of large investors that are highly specialized, and often dependent of outside financing which exposes them to agency issues The money comes from wealthy individuals, banks, companies, and other investors with only a limited knowledge of individual markets, and they may choose to liquidate their funds if the arbitrageur is not showing profit. In models without agency problems, arbitrageurs are generally more aggressive when prices move further from fundamental values because this raises the expected profit (Grossman and Miller 1988; DeLong et al. 1990; Campbell and Kyle 1993). When arbitrageur manages outside capital, however, his clients may infer from his short term loss that he is incompetent and, in lack of full comprehension of the strategy, withdraw their funding forcing arbitrageur to liquidate prematurely and realize losses. Fear of this scenario makes arbitrage less effective especially in extreme circumstances, where prices are significantly out of line and arbitrageurs are fully invested.

Most arbitrage involves several risks Shleifer and Vishny (1997); Barberis and Thaler (2003) separate the following: **(1)** fundamental risk because the long and short positions are not perfectly matched; **(2)** noise trader risk because mispricing can get larger and bankrupt an arbitrageur before the mispricing closes; and **(3)** implementation costs (e.g. short-selling costs – sometimes short-selling is simply impossible). Hence, these limits of arbitrage may prevent the markets from correcting prices that differ from fundamentals.

This means that combined investors' biases could affect market prices also in long term. Arbitrageurs as a group are also likely to disagree amongst themselves about fundamental value (i.e. they have heterogeneous expectations), and this could increase the general uncertainty they perceive about profitable opportunities, even in the long term. The smart money may, therefore, have difficulty in identifying any mispricing in the market. Hence, since funds usually are limited and investment horizons are finite, it is possible that profitable risky arbitrage opportunities can persist in the market for some time. The arbitrageurs may know that ultimately the prices will converge with the fundamentals and that they will make profit but if they have to borrow cash or securities to implement their position, pay periodic fees and report their profit position to their investors, they may be unwilling to take positions that will be profitable at some point of time but include too much exposure to market sentiment and possibility to run out of funds and/or time before that realizes (Cuthbertson and Nitzsche 2004, p.427).

The fact that arbitrage is in real markets limited is central to the behavioral finance's propositioning that market prices do not stay in line with fundamentals. The conditions for efficiency essentially require the average investor error to be zero, so that the investor bias is unsystematic, and that the covariance between investor error and investor wealth is zero so that any errors are not concentrated within the investor population (Shefrin and Statman 1994). In contrast to the neoclassical perspective, the behavioral perspective holds that informed traders (investors who are free from error) will not guarantee market efficiency, because informed investors are typically reluctant to take on the risk associated with the large positions required to eliminate inefficiencies completely. Therefore, the actions of noise traders (i.e. traders with biased beliefs, not based on fundamental information) may cause prices to be inefficient. As a result, arbitrage can be risky (Shleifer 2000). Mispricing, like when company value is less than its subsidiary, has been focused and verified in many studies, e.g. Cornell and Liu (2001), Schill and Zhou (2001) and Mitchell et al. (2002). Gromb and Vayanos (2010) present evidence of limits to arbitrage by studying cases where fundamental risk is practically non-existent, and arbitrage still is not capable of correcting price discrepancy.

Their examples fall into three categories: **(1)** stocks, with claims to almost identical dividend streams, can trade at significantly different prices; **(2)** stocks of a parent and a subsidiary company can trade at prices under which the remainder of the parent company's assets has negative value; and **(3)** newly issued bonds can trade at significantly higher prices than older bonds with almost identical payoffs. Thaler and Barberis (2002) and Scruggs (2007) argue that the one risk that remains is noise trader risk. ***"****Whatever investor sentiment is causing one share to be undervalued relative to the other could also cause that share to become even more undervalued in the short term".* These inefficiencies can also be long lived.

2.2.4 Value Stock Premium

According to the value effect, positive abnormal risk-adjusted returns accrue to portfolios of stocks possessing high ratios of fundamental values relative to their share prices. For example, high D/P, low P/B, cash flow to price (C/P). Thus, by examining the ratio of a stock's price (market value) relative to its fundamental value, stocks can be classified as either value stocks or growth stocks. Firms possessing high P/B, high C/P, high P/E and D/P are classified as growth stocks, while stocks at the other end of spectrum are classified as value stocks. Early tests on the value effect by Basu (1977, 1983) finds that companies with low P/E ratio earn positive abnormal returns. Further evidence of a value effect is reported by Litzenberger and Ramaswamy (1979) who document a positive relationship between dividend yield and common stock returns; Statman (1980) and Rosenberg et al. (1995) report a negative relation between P/B and average returns. Fama and French (1993) combine market beta, size and P/B in the equation for the cross-section of average equity returns and confirm the existence of the value premium.

There are number of possible psychological reasons why investors might overvalue growth stocks and undervalue value stocks: **(1)** investors may give too much weight to past performance, and end up overvaluing growth stocks and undervaluing value stocks; **(2)** investors may choose to invest in good companies because of herding and positive analyst coverage regardless that the price may already be too high. Lakonishok, Shleifer and Vishny (1994) argue that market participants irrationally extrapolate recent sales growth into the future and become overly optimistic about firms that grew fast in the past and overly pessimistic about firms with inferior performance, and hence end up overvaluing the former and undervalue the latter. Furthermore Shleifer and Vishny find no evidence that value stocks underperform growth stocks in bad states of the economy (recessions) hence rejecting the risk-based explanation for value premium. They state the covariation among value/growth stocks is probably caused by some other common factors than risk-based attributes. Kokkonen and Suominen (2011) suggest that time-varying level of mispricing could be such a property, and that misvaluations in the market can persist due to limits of arbitrage. Asness et al. (2009) agree and argue that the limits of arbitrage are a possible explanation of the prevalence of value and momentum effects in the markets.

2.3 Style Investing

2.3.1 Anomalies Based Styles

Traditional finance theory teaches investors to statically allocate money between assets but it is natural for human thinking to classify and group objects into categories based on some common characteristics among them. When making portfolio allocation decisions, many investors first categorize assets into broad classes such as large cap stocks, value stocks, government bonds, and real estate and private investments and then decide how to allocate their funds across these various asset classes (Bernstein 1995; Swensen 2000). On the other hand they can also be subdivided further into liquid versus illiquid, old versus new economy, domestic versus foreign, and combinations of each.

These categories are in modern investment vocabulary called styles. A style can also be a rule for stock selection like Joel Greenblatt's MF, discussed in this study, or Graham Dodd's/ Warren Buffet's stock selection screens. They are all published and have lots of investors applying them. The process where investors base their portfolio allocation on a style level rather than on an individual stock level is known as style investing (Barberis and Shleifer, 2003).

One advantage is that this greatly simplifies the decision process since allocating money across few asset styles is far less demanding task than choosing among thousands of listed securities. When anomalies for the traditional finance started to surface and become recognized in the 80s, also the finance industry responded with specialized investment funds using divisions in dimensions like small-large, value-winner, momentum-contrarian and so on. Many investors came aware that the different styles were more profitable in some time periods than others and started to use style rotation in anticipation to profit from this (Barberis and Shleifer 2003). Very good results have been achieved with investment styles that combine basic value strategies with some sorting and selection variables like e.g. in this study (enhanced value strategies).

2.3.2 Factor Based Styling

Finding the set of variables that best could forecast stock returns or alternatively classify stocks *ex ante* to winners and losers using advanced statistical methods, has become a subject of intensive research as availability and quality of data and software has greatly increased. Multi-factor asset pricing models that include factors based on firm characteristics such as size, P/B ratios, momentum (Carhart 1997), and liquidity (Pástor and Stambaugh 2003) became famous by the Fama and French (1993) where they added size and value factors to their model with a market portfolio to predict cross-section of stock returns with an impressive success and superseding the previously dominant asset pricing model CAPM. Since then an enormous amount of work that relates diverse groups of variables to either regression models predicting cross-section returns or to models that select stocks with screening type rule-based methods.. Examples of studies that use accounting based ratios and quantities to estimate the drivers of cross-sectional expected returns include: Chen et al. (2004), Gebhardt et al. (2001), Francis et al. (2003), Gode and Mohanram (2003).

Some of the most important papers on value drivers that has found significant in predicting stock returns include Ou and Penman (1989) and Piotroski (2000) who used several accounting based variables; Liu et al. (2002) used multiple variables that proxied valuation, growth and profitability, and suggest that using the forward earnings multiple yields higher valuation accuracy compared to using any historical multiples; Chan and Lakonishok 2004 linked P/D, P/CF and P/S to traditional value indicators like P/E and P/B with success; Easton (2004) and Bradshaw (2004) studied PEG (PE to growth) finding that an important predictor of future stock performance.

The factors included in the most elaborate style definitions, and regression techniques used to combine these factors, recognize that there are many dimensions to style that more simple methods (e.g. value, size, momentum) are not capable of capturing. These multifactor models can viewed as generalizing style concept Fabozzi (2003, p.49-51). For many financial forecasting problems, classification models work better than point prediction (level estimation); Leung et al. (2000) show this with stock market direction forecasting combined with trading rules for investment return maximization. Dutta et al. (2012) conclude that it is possible to predict out-performing shares by examining ratios calculated from financial reports. In practice, there are two approaches in applying a multifactor model: the statistical factoring approach and the fundamental-macro factor approach. To a large extent, the statistical approach is motivated by the APT theory of Ross (1976) which is discussed in the next subsection. In contrast, the fundamental–macro factor approach is a direct application of Merton's (1989) multifactor model which implies that any factors that influence the growth of consumption should also price individual assets.

In an early work by Marc Reinganum (1988) he was aiming in finding out if "winner" stocks tended to share certain common characteristics. He took a sample of 222 firms whose stocks had at least doubled in price during one year between 1970 and 1983. He then worked on to identify a set of common characteristics across the stocks in this sample and find out whether a successful trading strategy could be established on them. Reinganum identified a set of factors that were related to value, momentum, and size and enhanced value strategy with added ROI dimension. The results were impressive, outperforming the S&P 500 by 37 % (at a comparable risk level).

Following this approach, Robert Haugen and Nardin Baker investigated the predictive contribution of a large selection of factors grouped into five categories: risk, liquidity, price level, growth potential, and technical. Factor sensitivities were estimated using the 12 months prior to the beginning of 1993, and then expected returns for each stock were fitted for January 1993 using these sensitivities and each stock's exposure to these factors. Next, stocks were ranked from highest expected return to lowest and 10 deciles were formed, with decile 10 representing the 10% of stocks with the highest expected return down to decile 1 comprised of stocks with the lowest expected return. **Figure 3** presents the resulting returns for the decile portfolios.

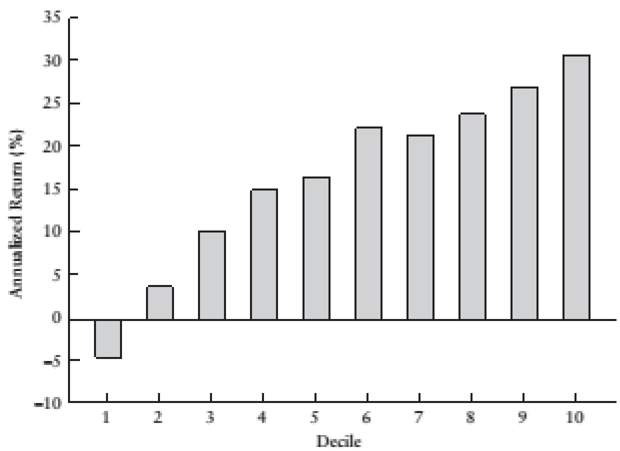


Figure 3: Average Decile Returns (Journal of Financial Economics, Vol 41, Issue 3, Haugen, R. A., and N. L. Baker, *"Commonality in the determinants of expected stock returns,* "pp. 401–39, July 1996.)

2.3.3 Enhanced Value Styles

While consensus of the superior returns on value stocks on average prevails among researches, portfolios of value stocks are often slow and uncertain to produce positive returns because large return dispersion across individual stocks. It turns out that the effectiveness of value investing relies on a small number of firms — using simple P/B investment criteria, less than 44% of low P/B firms earn positive market adjusted returns in the two years following portfolio formation, while the rest produce loss (Piotroski 2000). For that reason many researchers and practitioners have endeavored to enhance value strategies by including additional dimensions like capital return variables (Consistent Earner Strategy) and momentum factors (Recognized Value Strategy).

Many studies have shown that these enhanced value strategies are able to perform better in relation to risk and return than the whole market or one-dimensional traditional value strategies (Elze 2010). Financial statement analysis attempts to separate *ex post* winners from losers on the basis of information from financial statements that is not correctly represented in prices. Piotroski (2000) argues that such analyses will be especially effective in low P/B firms which are often ignored by market participants. By combining traditional value indicators with capital return (ROE), he indeed finds that financial statement analysis can effectively separate winners from losers with these methods. Piotroski's contribution is to use financial statement information to separate the good performers from the bad. Specifically, he uses nine fundamental signals to measure three areas of a firm's financial condition: profitability, financial leverage/liquidity, and operating efficiency. His screening method is capable of generating 23% excess return in years 1976-1996. Logical link between P/B ratio and ROE is that P/B can be interpreted as a measure of profitability expectations. When P/B ratio is below 1, the company is expected to destroy shareholder value. When the ratio is higher than 1, the company is expected to create shareholder value in the future. ROE, on the other hand, is a measure of profitability. Clubb and Naffi (2007), combining P/B with expectations and ROE, are able to explain significant amount of future stock returns.

Leivo and Pätäri (2011) combine value strategies with price momentum. Since value stocks may remain cheap for an extended period of time the value portfolio could be enhanced by including in the portfolio formation process an additional criterion that is used as a timing indicator for when to purchase value stocks. Their value-winner portfolio's average annual return over the 15-year sample period (1993-2008) would have been nearly 25 %, which exceeds the average stock market return by more than 10 percentage points. At the same time, the annualized volatility for the same strategy is 17.87 %, which is almost 4 % lower than the corresponding market volatility. As a result, the risk adjusted performance of this value-winner portfolio is significantly superior to market. Bird and Whitaker (2004) report that the added value attributable to each value and momentum strategy is basically uncorrelated, which enables performance improvement by combining these two investment strategies. In contrast, Bird and Casavecchia (2007a) report a significant outperformance of value-winner stocks against both the stock market and value-loser stocks when using price momentum as a sentiment indicator and S/P as a value indicator.

Leivo and Pätäri also argue that the Finnish stock market provides an interesting target market for this kind of analysis due to herding behavior of international institutional investors and relatively low liquidity. This also causes that individual investors are likely to be subject to more severe behavioral biases as it is harder to correctly value stocks. Daniel et al. (1998, 2001) and Hirshleifer (2001) show that greater uncertainty induces stronger behavioral biases. Therefore, opportunities to earn abnormal profits through active investment strategies might be better for two reasons. First, it is probable that pricing errors causing the value premium are larger in such conditions. Second, like results of Yu (2008) show the herding behavior of institutional investors stimulates price momentum. Elze (2010) studied multiple value strategies and found that including additional value indicators was not helpful, combining with ROE did improve significance with returns in level with one-dimensional value strategy and in combination with relative strength indicator (momentum) his best portfolio was able to produce 17.5% annual return premium to index. Aby et al. (2001) also found multiple value strategy not useful but that inclusion of ROE to selection criteria also improves investment return. Balvers and Wu (2006) show that combined momentum-contrarian strategies outperform both pure momentum and pure contrarian strategies.

All of the above mentioned studies of enhanced value strategies are examples of new but growing field of contextual fundamental analysis where commonly some variables (accounting and/or market variables) are used within some contextual group (e.g. style characteristics or industry group) to predict future stock performance. The intuition behind is that after grouping the stocks in relevant way separating winners/extreme performers from losers (i.e. finding the set of variables that can predict the future performance) will be much more effective than within wide stock universe since the subspace of stocks already share many attributes which can be controlled. Campbell et al. (2001) demonstrated that idiosyncratic volatilities have risen steadily while the market volatilities have remained stable. In relation, they observe, that stock returns fluctuate more closely among their peers that engage in similar lines of business or form some other company characteristics based groups. In other words, there is increasing heterogeneity across groups and increasing homogeneity among firms within a group.

Shen and Xu (2008) confirm these results and conclude: *"It is reasonable to assume that there exist unique factors that only influence stocks in a particular group in addition to common factors that determine the returns of all stocks in the market. A multifactor model with both common factors and unique factors could describe the return structure more efficiently and precisely than models with only common factors."*

Some pioneering texts of contextual fundamental analysis including already discussed Piotroski (2000), found value stocks a good candidate for contextual analysis; Trueman et al. (2000) studies variables that can classify internet stocks. They choose to analyze portals because that was relatively homogenous group with many shared characteristics; Beneish et al. (2001) studies extreme performers, both positive and negative, in order to be able to find set of predictor variables that could be used to separate them *ex ante*. They select the potential predictor variables from prior research and group them to four panels: firm characteristics, trading characteristics, market multiples and fundamental variables. Beneish et al. (2001) study is also important because it provides the actual reference framework to conduct contextual analysis. Sloan (2001) encapsulates the essence of contextual fundamental analysis as facilitation of the construction of more powerful models for explaining and forecasting company performance. The increase in predicting power comes from the introduction of new explanatory variables that are tailored to the particular context. This study builds firmly to these earlier papers using the companies that has been selected by GF as the context but, unlike them, applies DM methods to systematically find strong predictor variables. Hypothetically these companies share some group of attributes that are not necessarily all visible, nor easily identifiable, based on earlier common knowledge i.e. published papers. DM approach, in aim to discover new kind of information and dependency structures from the data, intuitively is very suitable for this kind of research problem..

2.4 Some Current Topics in Asset Pricing

2.4.1 Successors of Single Market Factor Asset Pricing

As an alternative to CAPM, Merton (1973) developed a multi-factor capital asset pricing model which allowed for a small number of state variables to affect excess returns of risky assets. The model was called intertemporal capital asset pricing model (ICAPM). It is a linear factor model with wealth and state variables that forecast changes in the distribution of future returns or income. In the ICAPM investors are solving lifetime consumption decisions when faced with more than one uncertainty. The main difference between ICAPM and standard CAPM is the additional state variables that acknowledge the fact that investors hedge against shortfalls in consumption or against changes in the future investment opportunity set.

Lucas (1978) and Breeden (1979) found that all relevant factors also affect consumption and hence designed consumption based CAPM (CCAPM). Alternatively, Cochrane (1991) and Balvers et al. (1990) proposed a production based CAPM. In addition, general equilibrium approaches have been suggested by, among others Cochrane (2000). The fundamental prediction of the CCAPM relates (with consumption beta) asset returns to their covariance with the intertemporal marginal rate of substitution (IMRS) which is the ratio between current and future marginal utilities of consumption of the representative investor. An individual asset is therefore more valuable if its return is expected to be high when consumption is expected to be low (when marginal utility is high). Thus, the systematic risk of an asset is determined by the covariance of the asset's return with respect to consumption (rather than its covariance with respect to the return on the market portfolio as in the CAPM). The CCAPM follows from the first order condition for an utility-maximizing agent's intertemporal consumption and investment choice problem. In equilibrium, the agent invests to the point where the marginal utility lost from foregoing current consumption is equal to the discounted expected marginal utility gained from that investment in the future.

An obvious extension to this simple formulation is to allow for habit persistence in consumption. Under habit persistence, an increase in current consumption lowers the marginal utility of consumption in the current period and increases it in the next period. Hence, the more consumer eats today; the hungrier he wakes up tomorrow as in Heaton (1995) and Constantinides (1990) or for time non-separability (utility of consumption in year 2 is not independent of the level in year 1) as in Abel's (1990). Despite its theoretical elegance, empirically the standard form CCAPM has failed and cannot explain asset returns (e.g. Campbell 1996). Possible explanations include faultiness of the preference model and incompleteness of markets with heterogonous agents instead of homogeneous as the theory assumes. There is, however, widely shared understanding that consumption based asset pricing models are the way of the future. The assumptions, functional forms, utility models and other included state variables are parts that will comply with more realistic structures. Habit formation theory discussed above is actually a bridge between behavioral and traditional economics. Cochrane (2005, p.126) argues in favor of the consumption based discount factor models by stating that: *"I think, the focus on means and variance, the mean-variance frontier and expected return/beta models is all due to an accident of history, that the early asset pricing theorists happened to put mean and variance on the axes rather than state contingent consumption."*

In the 1990s CCAPM extended to kernel pricing or the stochastic discount factor (SDF) framework models. SDF relates payoffs to market prices for all assets in the economy. This effectively is an application of the Arrow-Debreu model of general market equilibrium to financial markets. A state price exists for each state of nature and the market price of any financial asset is just the sum of its possible future payoffs (in each state) weighted by the state price. Making further assumptions about the economy can bring more specific results: assuming market completeness makes the SDF unique; assumption of linear relation SDF with macroeconomic shocks gives rise to linear model for asset returns; if the economic model is based on an assumption of a representative agent with an unambiguous utility function, then the SDF is related to the marginal utility of aggregate consumption. Also the nonstandard preferences and irrational expectations that inherently belong to modern behavioral financial analysis can be analyzed in this framework (Campbell 2000).

An SDF has the following property: value of a financial asset equals expected value of product of payoff on the asset and the SDF. An asset pricing model identifies a particular SDF that is a function of observable variables and model parameters. Hence, an SDF relates future cash flows of any financial instrument to their respective present market values. Most existing asset pricing methods can be shown to be particular versions of the SDF model. This includes the CAPM, CCAPM and even Black and Scholes option pricing theorem. Cochrane (2005) argues that there are two approaches to asset pricing: absolute and relative asset pricing. Absolute pricing involves pricing each asset with reference to its exposure to fundamental sources of macroeconomic risk. Relative asset pricing aims to price an asset with reference to prices of other assets. The SDF framework can be applied in both approaches. A key feature, not possessed by other factor pricing models, is that in the SDF model the asset returns are linear functions of the conditional covariances between the factors and the excess return on the risky asset. This is negative relation e.g. like in the CCAPM: the asset's return is higher and price lower the more negative the correlation with the SDF because asset's price is high when the marginal utility is low and *vice versa* (Smith and Wickens 2002). Equations **(1)-(5)** will clarify the above discussion on the SDF.

The price of an asset in period *t* is the expected discounted value of the asset's pay-off in period *t*+ *s* based on information available in period *t* (1).

$$P_t = E_t[M_{t+s}X_{t+s}] \qquad (1)$$

Where $\boldsymbol{P_t}$ = the price of the asset in period *t*, $\boldsymbol{X_{t+s}}$= the pay-off of the asset in period *t*+*s*, $\boldsymbol{M_{t+s}}$= the discount factor for period *t*+*s*, and $\boldsymbol{E_t}$= the expectation taken with respect to information available in period *t*. Thus $\boldsymbol{P_t}$ is the current value of the period *t*+*s* income$\boldsymbol{X_{t+s}}$. In general this income will not be known in period *t* and will be a random variable. The discount factor is sometimes called the pricing kernel, and it is a stochastic variable.

(1) Can also be written in terms of the asset's gross return which produces **(2)**.

$$1 = E_t\left[M_{t+1}\frac{X_{t+1}}{P_t}\right] = E_t[M_{t+1}R_{t+1}] \qquad (2)$$

Equations **(3)** and **(4)** clarify the relation between asset's return and the conditional covariance with the SDF ($\boldsymbol{M_{t+1}}$).

$$E_t(M_{t+1}R_{t+1}) = E_t(M_{t+1})E_t(R_{t+1}) + Cov_t(M_{t+1}R_{t+1}), \qquad (3)$$

$$E_t(R_{t+1}) = \frac{1 - Cov_t(M_{t+1}, R_{t+1})}{E_t(M_{t+1})}. \qquad (4)$$

Equation **(5)** shows how in CCAPM setting the SDF now takes the form of ratio of marginal utilities of consumption. $\boldsymbol{\beta}$ is the coefficient of time preference.

$$E_t\left[\frac{\beta U'(C_{t+1})}{U'(C_t)}R_{t+1}\right] = 1. \qquad \mathbf{(5)}$$

An alternative to the CAPM in determining the expected rate of return on individual stocks and on portfolios of stocks is the arbitrage pricing theory (APT) by Ross (1976). Broadly speaking, the APT implies that the return on a security can be broken down into an expected return and an unexpected component. For any individual stock, this unexpected component can be further broken down into component that is common to all stocks and specific that affects only this individual stock. The APT predicts that the common component will affect the rate of return on all stocks but by different amounts. For example, a 1% unexpected rise in interest rates might affect the return on stocks of a company that was highly geared more than that for a company that was less geared. So the risk of holding any security comes from two sources: **(1)** macroeconomic factors that affect all securities. These are factors that influence the whole asset market and cannot be diversified away. **(2)** company specific or idiosyncratic element. This element is unique to each security and, according to the APT, it can be diversified away. So in APT only the macroeconomic factors should be rewarded in portfolio return (Watsham & Parramore 1997). The APT, in one sense, is more general than the CAPM in that it allows a large number of factors to affect the rate of return on a particular security. The APT does not require any assumptions about utility theory or that the mean and variance of a portfolio are the only two elements in the investor's objective function. Since the nature and number of the priced factors are unspecified by the APT, two approaches have been used to empirically implement the theory. The most widely used approach, originally proposed by Gehr (1978) and subsequently extended by Roll & Ross (1980), relies on factor analysis techniques to simultaneously estimate the common factors and factor loadings of security returns. Also the principal component analysis can be used here, as a data reduction technique, to express the information in a large number of variables in terms of fewer derived dimensions. As applied to equity portfolios, an interpretation would be that these orthogonal combinations of observed variables are proxies for unobservable underlying economic differences in portfolios that are the cause of perceived portfolio characteristics. Viewing the factors in this way, they represent fundamental sources for differences in security returns (Coggin and Fabozzi 2003, p.87). The second approach is in contrast to the factor analysis approach. Chen et al. (1986) attempt to use macroeconomic variables to explain asset returns in the APT context. The macroeconomic variables are treated as factors in the APT return generating process.

The fact that asset prices tend to move intercorrelatedly suggest presence of underlying factors that are able to drive returns, but no one has yet determined which economic variables are responsible (Chen et al. 1986). Thus, there is no formal theory in choosing the appropriate group of economic factors to be included in the APT model (Azeez & Yonoezawa, 2003). Modigliani and Pogue (1988) suggest that the advantage of the APT is that it allows investors to specifically tailor their portfolios by adjusting the exposure to individual risk factors. This is contrary to the CAPM since the CAPM suggests that the market portfolio is the optimal risky portfolio, and that all investors will hold part or all of their investments in the market portfolio. Roll and Ross (1984, p. 24) also criticize that the market portfolio cannot possibly be optimal for everyone.

2.4.2 Realistic Risk Measures

Since the early days of MFT there has been a controversy about the concept of risk. Stoyanov et al. (2010) state that a reliable risk model should have realistic assumptions for return distributions and an appropriate downside risk measure. The MPT pioneered by the seminal work of Markowitz (1952) is based on the assumption that returns on assets are normally distributed. Under this assumption, total risk can be measured by volatility (variance of returns), and tail events deviating more than three standard deviations from the mean are very rare. Investors should hold mean-variance efficient portfolios because utility depends only on these two factors and hence it is only required to know the first two moments of the probability distribution in the considered asset's returns to create optimal portfolios. This approach would be correct if either investor's utility function is quadratic or the distribution of returns is conditionally normal. The first assumption requires that investors increase their absolute risk aversion as their wealth increase, while it is well known that they generally exhibit a higher risk aversion with lower levels of wealth. Financial returns are skewed and leptokurtic and thus higher moments are needed (Mandelbrot 1963). It has also been established that loss aversion has a substantial impact on what investors consider to be an efficient portfolio and that mean-variance analysis alone with the above mentioned utility models can be misguiding.

All dispersion measures are measures of uncertainty which is not necessarily risk because both positive and negative deviations from the mean are penalized. Also analysis based on semi-variance tends to produce better portfolios than those based on variance. This is why Markowitz favored the semi-variance of returns as a more appropriate risk measure, and it is discussed extensively in Markowitz (1959). A natural extension of semi-variance is the lower partial moment risk measure, also called downside risk or probability weighted function of deviations below a specified target return (Bava 1977; Fishburn 1977). This risk measure depends on two parameters: **(1)** a power index that is a proxy for investor's degree of risk aversion, and **(2)** target rate of return that is the return that must be earned at a minimum to accomplish funding of the plan within cost constraint. It works also considering expected moment of deviations above a specified target return, called upper partial moment; which can be used as a method to value excess return. Most people would prefer some loss limit, whereas a high standard deviation of profits is concerned a good thing. In particular, Karni (1985) have emphasized that investors' choices are strictly dependent on the possible states of the returns. Thus, investors have generally state dependent utility functions. Asymmetric risk preferences also became a subject in risk modeling with two main approaches: **(1)** incorporating skewness (and in some cases kurtosis) into the existing asset pricing models, and **(2)** estimating risk based on partial moments, mostly calculated from a portion of the return distribution that falls below a certain threshold (e.g. zero or risk-free rate). Ang et al. 2006 write: *"Kahneman and Tversky's (1979) loss aversion preferences and the axiomatic approach taken by Gul's (1991) disappointment aversion preferences allow agents to place greater weights on losses relative to gains in their utility functions. Hence, in equilibrium, agents who are averse to downside losses demand greater compensation, in the form of higher expected returns, for holding stocks with high downside risk"*

Many new risk measures take the empirically observed investor preferences into account and either measure only losses like Value at Risk (VaR) and conditional Value at Risk (CVaR), or take the asymmetries between profits and losses into account, such as the Omega risk measure that captures asset's idiosyncratic downside risk and upside potential (Gilli et al. 2006). Lately, the Omega ratio which was originally introduced by Keating and Shadwick (2002a) has become a dominant performance measure within the downside risk framework due its complete non-reliance on any distributional or utility model assumptions.

Keating and Shadwick argue that it performs better than Sharpe ratio even when returns are normal because it incorporates individual investor's arbitrary loss threshold into risk-return analysis. When returns are non-normal it is superior due to its design which includes information about all of the distribution moments of the underlying asset. The choice of utility function, level of risk aversion and level of loss aversion affects the investment process. These differences in preferences lead to different comprehensions of efficient portfolios, and also investors' portfolios will differ substantially in their stylized facts and properties (Winker et al. 2008, p. 8-10).

The VaR figure has two important characteristics: **(1)** it provides a common consistent measure of risk across different positions and risk factors; **(2)** it enables to measure risk associated with fixed-income position in a way that is comparable to and consistent with a measure of risk associated with equity positions. The other characteristic of VaR is that it takes account of correlations between different risk factors. If two risks offset each other, the VaR allows for this offset and tells that the overall risk is fairly low. If the same two risks do not offset each other, the VaR takes this into account and returns a higher risk estimate (Dowd 2002, p.10). VaR identifies a threshold, but not the magnitude of loss beyond the threshold. It only tells the largest loss in good states where a tail event does not occur; it tells nothing about what can happen in bad states where a tail event does occur (i.e. where loss in excess of the VaR). VaR's failure to consider tail losses can create counterproductive outcomes. For instance, if a prospective investment has high expected return but also involves possibility of a very high loss, VaR based decision calculation, which in simplest case is based on linear approximation of losses with normally distributed risk factors, might suggest that the investor should go ahead with the investment if the higher loss does not affect (and therefore exceeds) the VaR, regardless of the sizes of the higher expected return and possible higher losses. This is not traditionally considered prudent risk-return analysis, and can leave investor exposed to very high losses. VaR can also create moral hazard problems when traders or asset managers work to VaR defined risk targets or compensation packages. Traders who face a VaR defined risk target might have an incentive to sell out of the money options that lead to higher income in most states of the world without affecting the VaR and then occasional large hit when the firm is unlucky. VaR also lacks the most important property than a conservative risk measure should possess namely sub additivity. This tells that a portfolio made up of sub-portfolios will risk an amount which is no more than sum of the risks of constituent sub portfolios.

It reflects expectation that aggregating individual risks should not increase overall risk. VaR is therefore not a coherent risk measure as defined by Artzner et al. (1999) where he lists the so called axioms of coherency: **(1)** wealth at risk declines when added an amount of riskless wealth; **(2)** more wealth is preferred to less wealth; **(3)** aggregated risk of two investments is lower than the sum of the two associated single risks and **(4)** when the wealth at risk is multiplied by a positive factor, risk must grow also with same proportionality. These properties are mathematically formalized in axioms that ensure the coherency of a risk measure. In general, risk measures can be viewed as belonging to one or more categories, such as coherent (Artzner et al. 1999), consistent with respect to stochastic dominance – risk measure leads to same ordering of investment choices for all utility functions belonging to a same category (e.g. concave; increasing and concave), practical (easy to implement). Dhaene et al. (2003) classify risk measures based on whether they consider the entire set of outcomes, referred to as overall risk measures, or only the tails, the so called downside risk measures. An overall risk measure is a measure of the distance between risky situation and corresponding risk-free situation when both favorable and unfavorable discrepancies are taken into account. A downside risk measure is a measure of the distance between the risky situation and the corresponding risk-free situation when only unfavorable discrepancies contribute to the risk (de Vries et al. 2006).

CVaR, also known as Expected Shortfall (ES) or Expected Tail Loss (ETL) is an additional risk metric that estimates the magnitude of expected loss. ES is defined as conditional expectation of the return that falls below VaR (Yamai and Yoshiba, 2002). It is more modern and considered superior to VaR It is also coherent risk measure as described above. ES estimated at the 1 percent confidence limit is the expected return 1 percent of the time. ES depends on the shape of the distribution beyond the confidence limit. This risk measure depends on an additional parameter, the power index, which varies with investor's degree of risk aversion (Biglova et al. 2004). It has also received some criticism since it does not take full account of the severity of extremes, because it focuses on the mean shortfall. As a result of this, ES can give implausible rankings of relative riskiness and can fail to take full account of the impact of extreme losses (Wirch and Hardy 1999; Wang 2002). Another major problem with ES measure is that it treats a large probability of a small shortfall as equivalent to a small probability of a large shortfall. As argued earlier, however, investors tend to view losses differently from gains.

The ES measure has drawbacks if investors view consequences of large losses per unit differently from small losses. For example most people insure their houses, but do not insure many minor items that may have a higher probability of loss than the house. The mini-max risk measure was first used in portfolio selection problems by Young (1998). It represents the maximum loss over all past observations. It is possible to show that the mini-max risk measure is an extreme and special case of the CVaR. Therefore, the mini-max risk measure satisfies all the properties of the expected tail loss. Marginal VaR is the sensitivity of VaR of a portfolio to addition of a new position. It is the difference between VaR of the existing portfolio and that of the new portfolio, and it provides means of determining whether the potential addition moves portfolio closer to the efficient frontier. For example, two candidates for addition to the portfolio might have identical expected returns, but if the first is positively correlated with the old portfolio (increases VaR) while the second is negatively correlated (decreases VaR), the second choice is the better one (Ray 2010, p.35-40).

Apart from the full distribution modeling, when dealing with extreme (very low probability, very high impact) events, another method for modeling extreme events exists. It is based on extreme value theory (EVT) originally developed to model extreme events in nature and provides a model only for the tail events of the distribution; other methods need to be used to handle the rest of the probability mass. EVT is concerned with determining the asymptotic limits that describe the distribution of extremes. For example, EVT might be used to estimate the distribution of the maximum wave height (not the entire distribution of wave heights) based on only the maximum observed wave heights from the previous 100 years. Use of EVT to estimate VaR is appropriate when, as in finance most of the time, the distribution of returns has been observed to have fat tails. If the market dynamic that generated fat tails is unknown, then extreme value VaR (EV VaR) might be best estimated by fitting a curve to the observations beyond some quantile. Use of EV VaR is most appropriate when there is substantial tail risk (Dowd and Blake 2006)

2.4.3 Behavioral Asset Pricing

Modern behavioral and neoclassical asset pricing theory is built around the concept of a SDF (Cochrane 2005; Shefrin 2005). The concept is very flexible, and allows most asset pricing models to be expressed as special cases of a general framework. The neoclassical SDF is structured as a monotone decreasing function, often interpreted as the marginal rate of substitution for a representative investor. In log-log space, the negative of the slope of the SDF is often interpreted as the representative investor's coefficient of relative risk aversion. In many neoclassical models, the SDF is treated as time invariant. However, some authors use models displaying time variation in the SDF. Examples are Constantinides (1990) and Campbell and Cochrane (1999) which feature stochastic risk aversion, the result of habit formation. Habit formation is determined with reference consumption levels, and is therefore related to reference point-based behavioral preference models (e.g. PT). PT was one of the first models for decision under risk that permitted deviations from rationality and achieved theoretical tractability at the same time. It suffered from some deficiencies in its method of transforming probabilities which led to violations in stochastic dominance (Fishburn 1978; Kahneman and Tversky 1979, p. 283-284). The problem has been solved by Quiggin's (1981, 1982) rank-dependent expected utility (RDEU). Tversky and Kahneman built on Quiggin's idea and combined the descriptive advantages of original PT with the theoretical advantages of RDEU. Their cumulative prospect theory (CPT, 1992) provides now the most important non-expected utility model presently available. An additional advantage of CPT as compared to original PT is that it can also be applied to uncertainty i.e. situations when probabilities of events are not given (Chateneuf and Wakker 1999). In CPT like in RDEU a probability weighting function is used to transform outcome probabilities into subjective decision weights but CPT additionally distinguishes between gains and losses relative to a reference point. Probability weighting function allows capturing the emphasis placed on probabilities of events separately from their associated utility levels. One commonly used probability weighting is the inverse S-shaped which over-weights probabilities in the left and right tails of the outcomes distribution. Empirical support for this shape is based on a large number of experimental studies in economics and psychology. For example, Camerer and Ho (1994) provide evidence from several studies supporting nonlinearity of probability weights with higher sensitivity of preferences to tail events.

There are two major points of distinction between that distinguishes the behavioral and traditional approach: first is the concept of sentiment and second is assumption of EU. Traditional asset pricing theorists assume that investors seek to maximize EU which is rationality based framework. Proponents of behavioral asset pricing suggest that people generally behave in ways that are inconsistent with EUT and instead behave more in accordance with a psychologically based theory, such as PT. A well-defined concept of sentiment is central in a behavioral version of asset pricing theory. Sentiment, which originates from systematic errors committed by investors, is treated as an important determinant of market prices. The definition of sentiment can be stated as percentage error in the expected return probability of an asset both at individual investor and market level. Sentiment is involved in pricing of all assets through the SDF. Log-SDF is the sum of market sentiment and a fundamental component (a linear combination of some factors) that serves as the neoclassical SDF (Shefrin 2005). **Figure 4** displays a traditional neoclassical SDF based on fundamentals alone with a behavioral SDF that reflects the sentiment function which is the difference between the two functions. If the mean investor error is zero, so that errors are unsystematic across the investor population, and the error-wealth covariance is zero, so that errors are uniformly distributed across the investor population, then market sentiment will be zero. That condition effectively describes when the market will be efficient in the short run (Shefrin 2007).

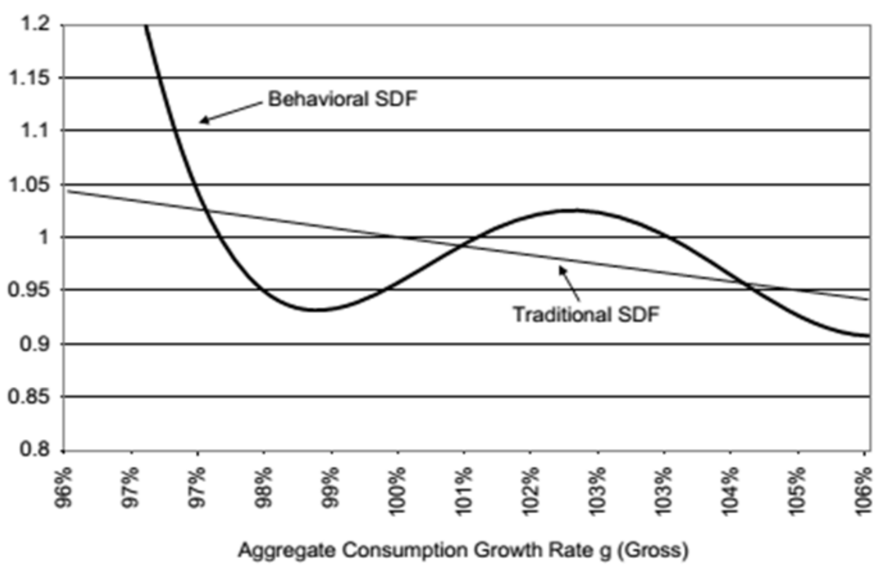


Figure 4: Behavioral and Traditional SDF (Shefrin 2008, p.4)

Tversky and Kahneman's CPT with rank-dependent probabilities has received the most recent attention among models that has been built as alternatives to EUT and to accommodate for its empirical inconsistencies. CPT has three principal components: a value function defined over (monetary) gains, similar to the utility function in EUT, a loss aversion function that transforms utilities over gains into utilities over corresponding losses (this function allows individuals to be risk averse over gains but risk seeking over losses, and for losses to matter more than gains), a weighting function used to transform probability distributions (allows the model to accommodate some violations of EUT).

In the context of financial asset allocation, the key elements of CPT are:

- Investors evaluate assets in comparison with certain benchmarks, rather than on final wealth positions.

- Investors behave differently on gains and on losses; they are not uniformly risk averse and are distinctively more sensitive to losses than to gains (the latter is a behavior called loss aversion).

- Investors tend to overweight small probabilities and underweight large probabilities. These elements translate respectively into the following technical features for the formulation of a portfolio choice model:

  - A reference point (neutral outcome – benchmark – breakeven point) in wealth that defines gains and losses.

  - A value function (which replaces the notion of utility function) that is concave for gains and convex for losses (such a function is called S-shaped), and it is steeper for losses than for gains.

  - A probability weighting function that is a nonlinear transformation of probability measure, which inflates a small probability and deflates a large probability.

Behavioralizing of the asset pricing theory is largely an effort of including behavioral preferences and heuristics into traditional models and improving their performance with added realism. **Figure 5** outlines the theoretical development and the root of asset pricing in short.

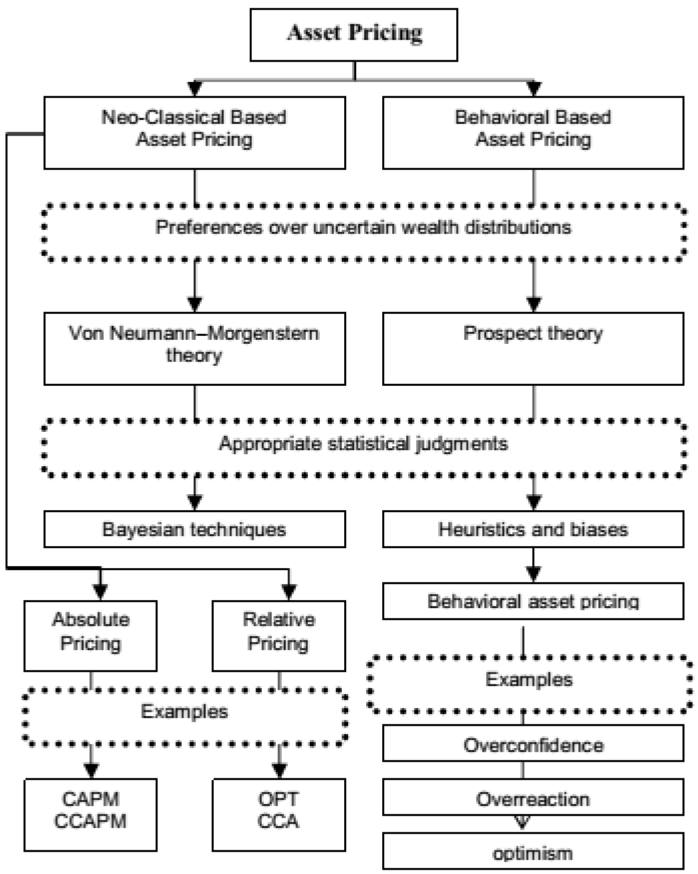


Figure 5: Frame of asset pricing standpoints (Çelik 2012)

The main distinction starts with the concept that how individual preferences over the distribution of uncertain wealth are taken place. PT is the behavioral counterpart for traditionalists' Von Neumann-Morgenstern utilities and Bayesian techniques are replaced by heuristics and biases. Cochrane (2005) claims that asset pricing problems are solved by considering how much absolute and how much relative pricing is appropriate, depending on the assets in question and the purpose of the calculation. Almost no problems are solved by the pure extremes. For example, the CAPM and its successor factor models are examples of the absolute approach. But in applications, they price assets' relative to the market or other risk factors, without answering what determines the market or factor risk premia and betas. On the relative side option pricing theory (OPT) and its generalization contingent claim analysis (CCA), involve assumptions beyond pure lack of arbitrage, assumptions about equilibrium market prices of risk.

3. METHODS AND DATA

In this chapter first a detailed description of the Greenblatt's Formula and the philosophy behind it is provided in section 3.1; next in 3.2 is an introduction to DM in relevant extent for this study; 3.3 documents the research process with the technical details and 3.4 presents some descriptive statistics of the data used in this study.

3.1 Joel Greenblatt's Stock Selection Method

3.1.1 Greenblatt's Investment Philosophy

Greenblatt's investment philosophy began to form in college where he didn't find EMH an intuitively appealing theory since markets seemed to work differently in reality. In a Foreword for book "*Market Sense and Nonsense"* (Schwager 2013), he writes: *" There, I was learning things about the efficient market theory (things that are still taught in MBA school to this day) that made absolutely no sense to me"*. On the other hand when he read about Graham's theories about value investing, he got interested and dropped out from law school to pursue a career in finance.

He describes this in an interview for a book by Jack Schwager (2012, p.454-484): "*A light bulb went off, and I started to read everything I could find on Ben Graham"*. His advices to stock investors are presented in his 1997 book *"You can be a stock market genius".* Its main messages are: **(1)** not trusting analysts' recommendations but doing your own analysis because that is the only way to find truly good bargains; **(2)** individual investor should use market timing and not diversify extensively; **(3)** historical volatility is not a good measure for riskiness; **(4)** investing in index fund can be a good alternative for someone who is not willing to put any effort in understanding the investment process but it generally cuts the profits that could be earned with minimal research and involvement; **(5)** value investing with rules like Graham's have proven to be superior to any other investment philosophies. He also warns not to trust the academics that say that the markets are efficient and there is no way to beat the index. Analysts' recommendations are often biased due to conflicts of interest like it is more profitable for investment bank to issue buy recommendations for a stock, and sell recommendations could cause loss of business fees and/or analysts' possibilities of meeting the management on part of the target company. It is also much easier for an analyst, who is career oriented, to be in line with the majority when calculating result and key ratio forecasts. The timing means here not to be necessarily in the markets all the time but to pick the spots where the sentiment is either very high or very low to trade, so that the market values deviate most from the intrinsic values. Volatility is not risk since stock that moves up significantly is not riskier than the stock that moves down slightly, and historical volatility extrapolated to the future doesn't tell about the possible hidden risks in the investment that realize only seldom. Investors should use measures like margin of safety which is the difference between the market value and intrinsic value and down side risk. About value investing Greenblatt makes three critical points: **(1)** value investing works; **(2)** value investing doesn't work all the time; **(3)** (2) is one reason why (1) is true. Buying good shares when they are cheap works in a long run, and the fact that it is not working all the time makes it more effective for those who can wait. To perform better than an average investor, one has to make contrarian bets, and there will be periods that this does not seem to work very well. In fact, he answers to questions about whether his system will work in the future, now that it is public, that it will probably work better since markets have institutionalized quickly, and professional money managers cannot wait results for a long time as the assets move soon to the last period's best performer. Greenblatt says that: *"Even if a manager knows that he should be looking longer term, his investors pressure him for performance over the near term"*.

3.1.2 Calculation of Greenblatt's Formula

Greenblatt's next book *"The little book that beats the market"* argues that there is even easier way to make money with stock investments that requires very little effort – *Buy good companies when they are cheap*. His formula accomplishes this by ranking stocks based on two factors*: return on capital* and *earnings yield*.

Particularly for this purpose they are calculated with following formulas:

Return on Capital = EBIT/ (Net Working Capital + Net Fixed Assets) **(6)**

In **(6)** EBIT (earnings before interest and taxes) is the pre-tax operating earnings and the denominator is *tangible capital employed*. There are several reasons not to use the more the more common ROE or ROA in this context. EBIT allows comparing companies from different branches and countries without distortions from variations in debt levels or tax rates. The purpose is to compare for each company the actual earnings from operations to the cost of assets used to, and actually needed, to produce them. This differs from the total assets in ROA (return on total assets) or equity in ROE (return on equity capital). Net working capital less excess cash is used because the company has to fund its receivables and inventory but payables are effectively an interest free loan. Short-term interest bearing debt is also excluded from current liabilities. In addition to working capital, a company must also acquire funds for the purchase of fixed assets needed to conduct its business. Tangible capital employed is thus calculated by adding net working capital with a deprecated net cost of fixed assets. Intangible assets, especially goodwill, are excluded because in most cases it is historical cost that does not need to be constantly replaced like tangible capital which therefore is more accurate reflection of a business's return on capital going forward.

Earnings Yield= EBIT/Enterprise Value **(7)**

In **(7**) enterprise value is calculated as *market value of equity + net interest-bearing debt*. Earnings yield is superior to e.g. P/E because enterprise value takes into account both the price paid for an equity stake in a business as well as the debt financing used by a company to help generate operating earnings. It is thus pre-tax earnings yield on the full purchase price of a business which again is comparable across different companies (except for banks and insurance companies which are excluded from the stock universe). Trailing 12 month historical values are used for earnings statement items like EBIT and latest financial interim values for balance sheet items and latest price quotation for market cap.

The next step in GF calculation is to rank the stocks that belong to the investing universe (e.g. stocks listed in the Nordic Exchanges as in here) based on their Return on Capital. The stock with the highest value will be assigned rank 1 and so forth down the list. Same is done for Earnings Yield, and then the both ranks are added together to come up with a total points number for each individual stock. Then the stocks are ranked again from lowest total points to highest. These (in the top) stocks are now the ones that have the best combination of both variables and are hence good and cheap with the chosen metrics. In an interview referenced above Greenblatt suggested investing even to first 20-30 stocks for those that are complete novices to stock investing but hinted that with some ability, better selection methods could be developed, and that in fact his team has put considerable effort to this and come up with further refinements for the formula.

3.2 Data Mining Methods in this Study

3.2.1 DM History

Data mining is usually referred as application of statistical, data analytical and machine learning methods to large data sets. Witten and Frank (2005, p. 8) define it concisely as a process of discovering useful patterns, automatically or semiautomatically, in large quantities of data.

DM is sometimes claimed to be a new discipline but in fact the term is mentioned in scientific journals already in early 1980s. Besides many of its core techniques are inherited from much older periods. The search for utilizable patterns from data has continued already centuries with methods like correspondence analysis, discriminant analysis, and logistic regression or the Bayes' theorem originating from the 1700s. The term has also been used in a negative context for over exhausting the data by testing a multitude of variables without any *a priori* hypothetical reasoning of correlation or causality until some combination fits to the data. Foster et al. (1997) provide theoretical analyses of data exhausting in the context of return predictability. He states that it is typically believed that out-of-sample tests provide protection against this as long as the test observations are not used in the model estimation. Apart from these concerns, the use of DM techniques in discovering completely new information and dimensions from the data should be completely acceptable practice in science. Traditionally these methods have been applied in e.g. credit scoring in banking business or customer scoring in database marketing in context of planning campaign offers. As an example of more contemporary uses are mentioned (Tufféry 2011, 2) pharmaceuticals industry that uses data mining in screening effects of chemicals and molecules to different diseases e.g. cancer. They may not know the exact effect mechanism beforehand but they get valuable ideas where to start developing medicines. Recently, automated inference with model and variable selection algorithms has raised great enthusiasm in empirical econometrics. Phillips (2005) discusses automated discovery in science and claims that advances in computer power, electronic communication, and data collection processes have changed empirical economics profession, elevated its status and opened new possibilities. Particularly, he emphasizes the ability to build econometric models in an automated way according to an algorithm of decision rules. Thousands of regressions and model evaluations may be performed in seconds, statistical inference may be automated according to the properties of the data, and policy decisions can be made and adjusted in real time with the arrival of new data. Empirical modelers are widely adopting the use of modern computing power and tailored software to search systematically for models with superior performance. Phillips also makes a point about the important challenge that the researches face in incorporating economic thinking and methods into the automated model and variable selection process.

Pesaran and Timmermann (2005) employ automated methods in real time econometrics for the use of businesses, governments, central banks and traders in financial markets with the focus on making decisions in real time, and who hence have an urgent need to develop robust interactive systems that use econometric models to guide the real-time decisions. They also suggest procedures to mitigate the differences in statistical inference with the traditional approach. Campos et al. (2005) argue in favor of this new empirical research paradigm against the conventional. In next quotation, in short, they say that for practical work in econometrics, the use of data driven methods is essential in today's world. This same idea applies to many other areas and data mining has thus become a main stream discipline.

> *"The economy is a complicated, dynamic, nonlinear, simultaneous, high-dimensional, and evolving entity; social systems alter over time; laws change; and technological innovations occur. Thus, the target is not only a moving one; it behaves in a distinctly non-stationary manner, both evolving over time and being subject to sudden and unanticipated shifts. Economic theories are highly abstract and simplified; and they also change over time, with conflicting rival explanations sometimes coexisting. The data evidence is tarnished: economic magnitudes are inaccurately measured and subject to substantive revisions, and many important variables are not even observable. The "conventional" approach insists on a complete theoretical model of the phenomena of interest prior to data analysis, leaving the empirical evidence as little more than quantitative clothing. Unfortunately, the complexity and non-stationarity of economies makes it improbable than anyone–however brilliant–could deduce apriori the multitude of quantitative equations characterizing the behavior of millions of disparate and competing agents. Without a radical change in the discipline's methodology, empirical progress seems doomed to remain slow."*

DM methods can be subdivided in two distinct classes, predictive and descriptive, which are also called supervised and unsupervised respectively. In unsupervised methods no target variable is identified for the DM algorithm but the patterns and structures among all the variables are searched. These are often used in dimension reduction in the data like clustering and principal components analysis (PCA). The data is described in new dimensions. On the other hand in the supervised methods **(1)** there is a particular target variable, and **(2)** the algorithm is trained with the data to adjust the model parameters for the best predictive properties. Aim is either to predict the level of the target or classify it to some predefined category. The most important are decision trees and neural networks but also classical models like logistic and linear regression.

The division is illustrated below in **Figure 6** with just a couple of examples of large pool of methods.

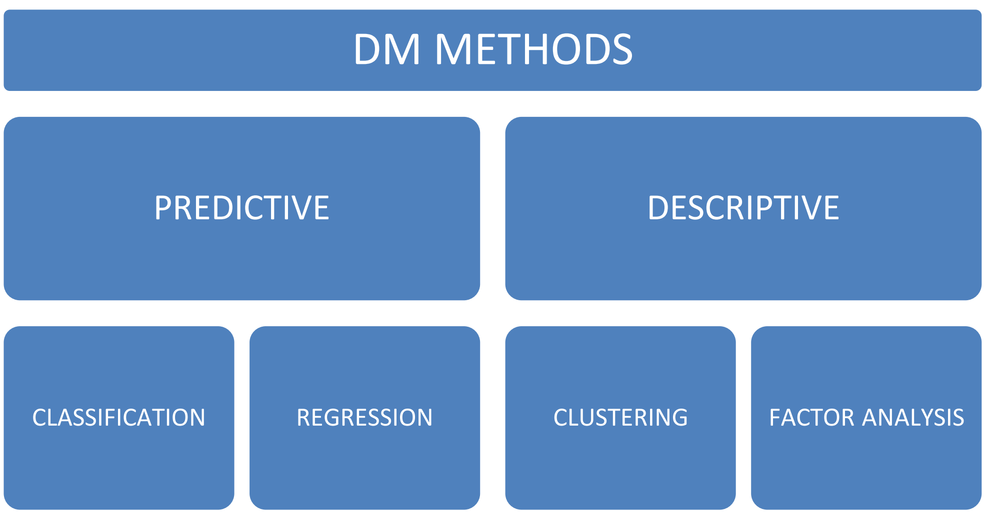


Figure 6: DM Methods

3.2.2 DM Process

There exist multitudes of different specifications of DM process. In common case the starting point is some electronic data storage: database, datamart or data warehouse, depending on the task scope. The version presented in this text is applied to the task at hand. **Figure 7** presents a quite generic operation flowchart. The link from end to beginning is important because when new data accumulates, also new questions may arise and process could restart with new or refined objectives. In practical analysis there could be a link from every phase box back to Problem Definition phase since this is also a learning process and when knowledge of the data accumulates the researcher also acquires a better understanding of the most achieving ways of working with it.

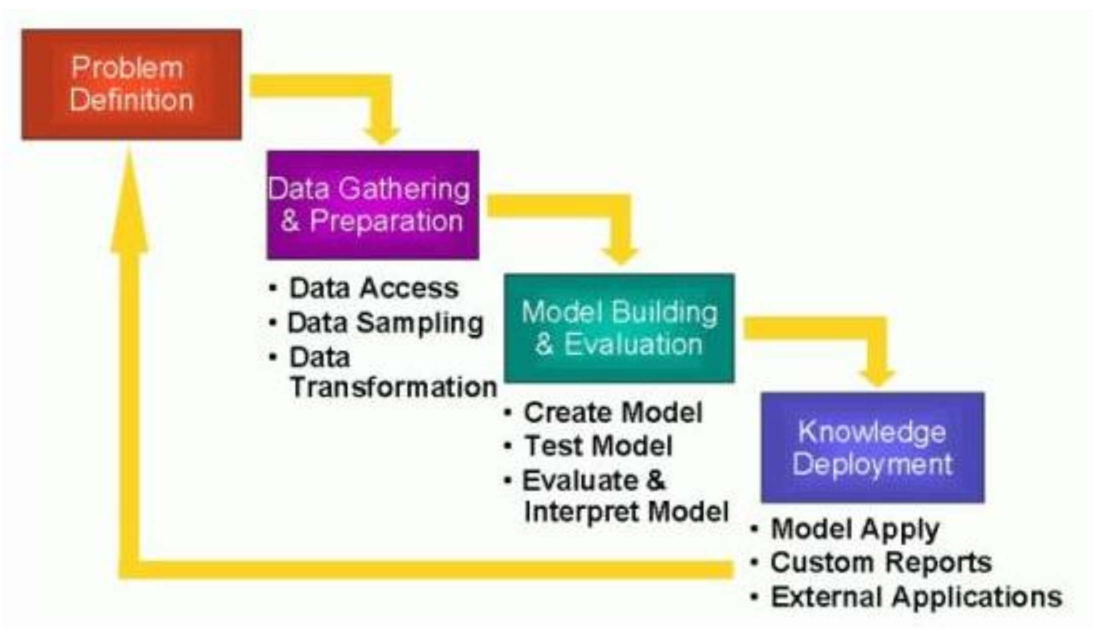


Figure 7: DM Process **(Oracle® Data Mining Concepts)**

Giudici and Figini (2009, p.1-4) list the tasks related to each process phase. The following list is an adaptation from their book with comments relevant to this particular project.

1. Definition of objectives

    - It is important to crystallize the goals of the DM project since this, to large extent, determines the methods that are applicable. It is not always easy to define the analyzed phenomenon statistically so this phase is one of the most difficult ones.

    - In this study the primary objective was to find a filtering method for the unsuccessful investments. Secondary goals were to separate also the highest returning stocks from the moderate risers. Also intuitive functional form with testable coefficients was considered important. The last criterion sets multiple or ordinal logistic regression as preferred to tree based classification algorithms and neural networks but they all still should get tested for performance in primary and secondary objectives. A probit model could be an alternative to logit but it requires more normally distributed data.

2. Selection, organization and pre-treatment of the data (data cleaning)

    - After the analysis objectives are clear, it is time to identify the available data sources and collect or select the variables for the initial data matrix. There are usually internal and external sources both openly accessed and proprietary that contain relevant information for analysis.

- The data needs to be quality controlled before the analysis: some variables may not be suitable or have missing/unreliable data. When some variable has part of the data missing, the analyst needs to carefully study and model it. The distribution of the cases with missing data in terms of the other variables is of interest, and it should be as random. as possible. Based on this study there is a decision to either delete the variable or choose an imputation (patching) method for the missing values. Otherwise the results may be biased.

- Internal data source in this study was the information produced by the investment simulation itself since it was obvious that variables like stock's rankings in both Earnings Yield and Capital Return as well as in combined scoring and an indicator for the stock's previous selections were interesting. As outside sources several databases for company and macro-economic data was utilized. After the data matrix was acquired SPSS Statistics® was then used to handle the missing data.

3. Data screening, transformation and exploratory analysis

- In this phase the data is screened visually to establish the distribution of the individual variables and assess the need for transformations. For many analysis methods it is valuable that the data is close to normal, and thus e.g. log transformation often improves the quality. Here also the possible anomalous data points, that are different from the rest, are detected. Outliers can have significant influence on the analysis outcome so it is important to consider carefully whether the point is erroneous or does it instead contain valuable information. The data needs also to be screened for multinomial outliers (a combination of variables that is unusual) that are not detected in visual inspection but by calculating the Mahalanobis distances which is a multivariate data point's relative distance in from a common midpoint. In this phase a need for additional data may be noticed and observations here often influence also the method choices in the next phase.

- Initial variable choices are made based on their power of influence on the dependent variable. Histograms, box and scatter plots are important visual aids in data screening but also pairwise and multivariate charts and tables.

- Most statistical packages have automatic tools for the tasks in this phase e.g. to screen and rank the variables for influence on the dependent variable or to handle outliers. They may help in large projects but should be used with caution. The automatic outlier handling routine transforms observations that lie beyond a preset number of standard deviations from the center to the border. If this is done routinely without consideration it could easily destroy valuable information. Variables' influence also often varies by the method and combination of other explanatory variables.

4. Specification of analysis methods and techniques

- There are numerous statistical methods available, and the decision depends on the information from the phases 1 and 3. The DM process depends on the project; its goals and data. The goals determine if the analysis's purpose is to describe the data or predict. This knowledge guides to select the relevant analysis method from one of the main groups in **figure 6**. After screening the data, the measuring scales (e.g. ratio, categorical, ordinal) and distributions (e.g. continuous, discrete, multinomial, binomial, categorical) of the variables are known which aids the analyst with the final model selection as some methods have requirements and performance differences on quality of data inputs. It is customary in DM to create several models using different algorithms in modeling the problem and then test and rank those models to end up with the best solution.

- The model and variable selection is an iterative process. Different methods will require a distinctive set of variables for optimal performance. Also some algorithms may work better with a certain types of data transformations. Here some automated modeling tools may assist the process by testing with dispatch the model with a number of variables and transformations of variables in order to find the most suitable combinations.

- It is important that when testing the models the test data comes from a partition that has not been involved in the model specification but from hold-out-sample to avoid data snooping i.e. to eventually, after large number of tries, come up with a combination of variables that appears to work well but in reality is not capable to predict the independent variable outside the fitting sample.

5. Evaluation and choice of the final model

- The model candidates are tested and ranked according to their performance with some predefined criteria. The desired feature can be e.g. the model's ability to predict correct classes as a total percentage or weight right and wrong predictions differently between classes. The target could be to find a model that can best separate rising stocks from falling ones and hence the weighting in that score would be heavier than on the ability to separate very high gainers from moderate ones.

- When the final model is determined and found fulfilling the analysis requirements, it is then deployed to production environment which in the case of this study could be a server with on-line access to data feed (e.g. Bloomberg, Reuters) and order placement facility for relevant stock exchanges. With this kind of setting an automated trading program could easily be created. At least SPSS Modeler® and some other DM software have out of the box functionality that supports this deployment scenario. Otherwise some programming language can be used to tailor the functionality suitably. After all it is the model's ability to produce some real, monetary, output that defines its usefulness.

3.3 The Research Process

3.3.1 The Research Framework

To test the GF's investment performance (later analysis stage 1), a computer program was made with using MS Excel® VBA® programming language. The program ranked the stocks in each quarter from 1Q2000 to 4Q2011 in the way described above. 5 highest ranking stocks in each quarter were chosen to the investment portfolio; they were held two quarters and then sold. This was empirically found to be optimal choice for the portfolio size and holding period. In Greenblatt's book the period is one year due to tax reasons but really there is no theoretical reason to stay with it. The portfolio size also is much larger in Greenblatt's experiment but the testing revealed that in the Nordic markets portfolio size above 8-10 stocks and holding period above six months would have severely impaired the investment return. HPR (holding period return) is recorded (including cash and stock dividends and coupon issues) where trading costs are taken into account. Taxes from transactions are paid in the end of the following year. The weight of one stock in portfolio could increase without a limit if it got selected in consecutive periods.

In GF like in other value investment strategies one problem is that some stocks do not perform well and some perform very badly with only a small selection of stocks that are extreme performers. The extreme performers can generate enough profit for the whole portfolio to yield extraordinary well. Obviously there is a strong motivation to be able to *ex ante* separate those extreme performers from other stocks. This problem is closely related with the other referenced papers in the field of contextual fundamental analysis (e.g. Piotroski 2000 and Beneish et al. 2001). In this study the approach is different from those earlier ones, and instead of relying factors that have worked well in other contexts, DM techniques are applied with IBM® SPSS® Modeler to systematically produce a model that can among these stock selections predict their future performance (later analysis stage 2). Three categories are used for the stock's HPR which in this case is the total return in 6 months following the selection: HPR<0 –> *loss*, 0 ≤ HPR< 30% –> *profit* and HPR ≥ 30% –> *strong profit*.

3.3.2 Model and Variable Selection

The essence of model selection is to maximize the model performance and do it parsimoniously (with as few parameter as possible). Adding parameters usually improves the predictive capacity but lost degrees of freedom impair the model effectiveness. Another problem in excessive complexity is overfitting where increasing the number of independent variables models noise in the training sample and leads to reduced performance in out-of-sample data (Tabachnick 2007, p.11). Often, also when the number of explanatory variables is increased, they become correlated with each other. This so called multicollinearity makes the model unstable with symptoms like erroneous signs in parameter estimates and high standard error in them. Typically the F-statistic may indicate that the model is strongly significant but none of the individual parameters reaches statistically significant *p*-value. It is common that this kind of model does not generalize well to out-of-sample data.

When the aims of the analysis have been stated sufficiently explicitly, together with the quality of the data, they will guide far in choosing appropriate model class and algorithm. In DM it is common to use automatic tools in aid of this process but they should not be given the last word. In variable selection, on the other hand, the purpose is to find variables that have strong influence on the modeled target without interest in the functional form. But in the later stage of the analysis they are aligned so that the particular set of variables is used with each possible functional form (Clarke et al. 2009, p.582). This separateness allows to first do screening for both with automated algorithms to narrow the universe of possible combinations. Finally potential models are tested and ranked to find the best one. This study aims to build a model for predictive classification so relevant model classes were decision tree, logistic regression and neural networks. **Table 2** lists some characteristics of them that should be taken into account. Variable and model selection tasks are iterative integrated processes where intuitively and automatically selected best variable set feeds back to model class selection and then back to variable set narrowing to final model. Clarke lists the choices analyst must make during the process: model class, used variables, functional form and correct parameter estimates.

Table 2: Model classes

Class	**Assumptions**	**Important characteristics**
Decision Tree	Doesn't require any assumptions for distribution.	- tolerates multicollinearity - propensity to overfit - can handle categorized and qualitative variables - produces interpretable model
Logistic Regression	No normality assumptions for data distribution.	- sensitive to multicollinearity and outliers - can handle continuous or categorized variables - parameters are interpretable
Neural Network	Doesn't require any assumptions for distribution.	- tolerates multicollinearity - no interpretable model

3.3.3 Model and Variable Selection in this Study

The original data matrix included 50 variables, extracted from various internal and external data sources, which were considered possible candidates for the predictive classification model. Three transformations of each were formed: categorized, standardized and original scale. Automatic selection algorithms were then used to narrow the universe of model candidates. Variables (including all transformations) were grouped and coded with letter and number codes s.t. letter A-C stands for the group, the number is for identification and the letter I = *interval scale*, N= *nominal scale*, O= *ordinal scale*, T=*standardized* marks the scale/transformation attribute of the variable. For example A1I is a continuous variable from group A and A1O and A1T are categorized and standardized transformations of the same variable. Group A includes historical financial ratios and variables derived from them.

These variables illustrate broadly the operational result, effectiveness and financial position of the company. As an example from this group is net debt / total equity. Group B consists of simulation model produced data e.g. earnings yield rank, and C is for macro-economic variables which include variables like principal components from matrix that has e.g. GDP growths of the Nordic countries in last 6 months as columns. PCA was applied because many variables in that group were highly correlated and were not significant on their own. In addition to individual variables the data was screened for 2-way interactions between them. Sometimes *new variables* formed this way are significant even though the constituent variables are not. Interpretation of these variables is similar to principal components or factors. They represent latent variables that can be proxied with linear combinations of measurable variables. Intuitive meaning should always be evident for the variables formed in this way to be useful in analysis.

The task was to find a model, using data set available at the investment time that is capable of predicting the HPR of the stocks selected by GF model simulation program in analysis stage 1. The stocks' HPRs were mapped to a three class ordinal scale (used as nominal scale) variable (3= *loss*, 2=*profit*, 1= *strong profit*), and the main emphasis was to be able to filter out the *loss* stocks but also the correct separation of the stocks in the *strong profit* category was of great interest. When the above model classes (in **Table 2**) were tested running combinations of independent variables with automatic variable selection algorithms, it became clear that multinomial logistic regression with categorized independent variables was the best performing alternative for model choice, and there is an added benefit of interpretable model parameters as parameters in logistic regression are directly related to category probabilities. It is also known that logistic regression with categorized variables performs better than with the original scale (Tufféry 2011, 77). Ordinal logistic regression was another possible choice. This model class would have had an additional benefit of more parsimonious model as only one set of function parameters needs to be estimated. The explanatory variables all have same coefficients in the functions for all of the categories. Only the location e.g. the constant term of the separation function changes between the sequential categories. Ordinal logistic regression requires a natural ordering of the categories and that they share all other ways similar attributes.

With this data the requirements for ordinal categories were not met, and instead two functions needed to be estimated. With the model class determined, the final variables were chosen using partly automated tools and forcing intuitively important variables and those that also had had significant predictive power in previous test runs into the selection process. In the final set all A-group variables are categorized and B and C-group in original scale. From the initial 150 variable data matrix (including all transformations) a 9 term model was designed (variables, interactions and a constant term) which was able to predict the correct response category in all partitions with 80%-90% accuracy as is explained in Chapter 4 with more detail.

3.3.4 Automated Variable Selection Algorithms

There are three main types of in DM often used automatic search algorithms for variable selection in regression analysis*:* *forward selection*, *backward elimination* and *stepwise*. Forward selection algorithm starts with no variables in model and adds in every round the one that improves the model performance most. Correspondingly backward algorithm starts with all selected variables in model and then eliminates sequentially until stopping criteria is met. Stepwise is bi-directional and in every step it checks variables that meet inclusion or exclusion criteria (e.g. Wald, LR). Variable selection algorithms should be used in conjunction with *best subsets* algorithm since they can easily skip the best model having the entries and the exits executed one by one, and as is well known, significance of the term is often conditional to the other terms in the model. For this reason the automatic tools should only be used as an aid in screening large number of variables. Intuition should be the first and the last guide in variable as well as model selection. This is why the analyst should force the variables that are intuitively important to stay in the process as long as he is absolutely convinced that a better model has been reached. Critics also state that models created with automated variable selection overestimate parameter significance (Sheather 2009, p. 238) due to multiple comparisons, and those *p*-values are hard to interpret since they are conditional on previous step tests. The parameter estimates should thus be biased.

Whether these claims really have scientifically sound justification – or not, is irrelevant in practical applications since an obvious solution is to divide the data into training, testing and in some cases validation partitions so the deficiencies in the fitted model will be revealed by the inferior forecasting ability when applied to out-of-sample data. Tufféry (2011, 85) also strongly advises to the use of automatic variable selection algorithms in assistance, especially with logistic regression models.

3.3.5 Logistic Regression

Linear regression is used to model continuous response variable (dependent variable) against one or more explanatory variables. Instead, when the response variable is categorical, must logistic regression (or one of a few other technics) be applied. Its functional form is not linear but sigmoid in variables but it is conveniently linear in parameters. Logistic regression function is the most important function in this class. **(8)** Presents logistic regression equation in one explanatory variable. For each category of response variable a new curve will be drawn as it represents odds relative to base category. If more explanatory variables are added to equation, β and x are vectors. **Figure 8** illustrates the shape of the curve. Tufféry (2011, 477) lists attractive features of logistic regression: **(1)** explanatory variables can be arbitrary scale; **(2)** response variable can be ordinal or nominal; **(3)** no distributional assumptions for explanatory variables are needed; **(4)** it provides very accurate models even with small samples; **(5)** the models are concise and easily programmable; **(6)** it allows for interactions in explanatory variables; **(7)** it models directly a probability and **(8)** it allows stepwise selection of variables. The DM process in this study is a good example of how the process goals and the data quality helps in determining the most suitable analysis method as all the above features fit into the model requirements stated for the 2nd stage analysis..

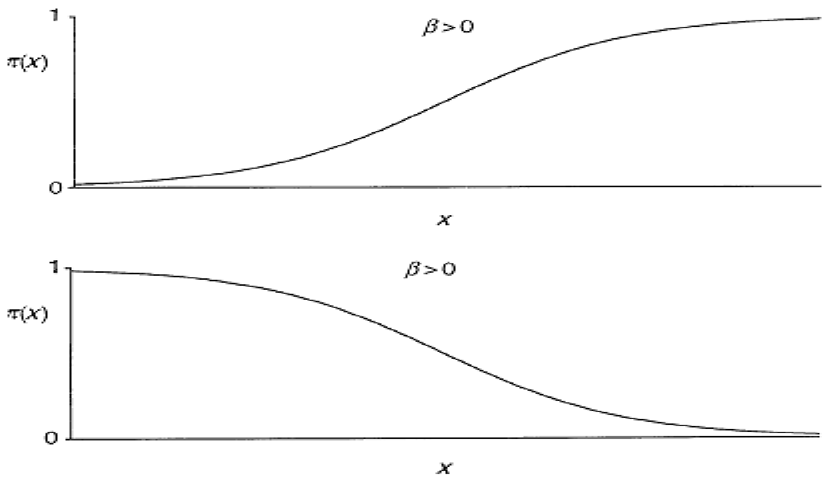


Figure 8: Logistic regression (Agresti 2007, 70)

(8)

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}$$

Assumption in logistic regression is that logs of odds ratios of response categories can be expressed in linear combination of independent variables (Zopounidis et al. 2008, p. 470). This is also the format from which its parameters are estimated **(9)**. Logit in itself is logarithm of odds ratios i.e. ratios of probabilities of event happening $\boldsymbol{\pi(x)}$ to that it do not happen 1-$\boldsymbol{\pi(x)}$. Sign and magnitude of parameter β determine direction and strength of effect of the independent variable.

(9)

$$\operatorname{logit}[\pi(x)] = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x$$

(10)

$$\frac{\pi(x)}{1-\pi(x)} = e^{\alpha+\beta x} = e^{\alpha}(e^{\beta})^{x}$$

(10) Provides an intuitive interpretation for the model parameters: odds for event shift β units when *x* changes by one. So this formula reveals easily the strength of explanatory power of (any) independent variable. When there are more than two response categories the model is called multinomial logistic regression **(11)** which has one less equations than the response variable has categories since one of them is chosen to be the base category to which others are compared with the odds ratios.

Classification results are indifferent to the choice of the base category but distinct set of equations is defined for each choice since the equations model the probabilities of the move from the base class to the target class. An alternate equation of the model **(12)** provides directly the probabilities of the response categories.

(11)

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j x, \quad j = 1, \ldots, J-1$$

(12)

$$\pi_j = \frac{e^{\alpha_j+\beta_j x}}{\sum_h e^{\alpha_h+\beta_h x}}, \quad j = 1, \ldots, J$$

So the probability of category *j* is the exponential of its equation divided by the sum of exponentials of the functions for the other categories. If *x* is vector-valued, the equations do not necessarily include all the same explanatory variables since $\boldsymbol{\beta_{hi}}$ may be zero for some$\boldsymbol{x_i}$.

3.4 Data

3.4.1 Data Sources

The company fundamental data for the first analysis stage, GF stock selection and investment study, was acquired from Thomson-Reuters's® Thomson Financial® database which had the best coverage among companies listed in the Nordic exchanges. Each listed company in Finland, Sweden, Norway and Denmark that had data available in the database, excluding banks and insurance companies, was joined with six data items: EBIT, Enterprise Value, Working Capital, PPE (property, plant and equipment) and end of period value for TotalReturnIndex calculated by Thomson Financial for each quarter from 1Q2000 to 4Q2011. The TotalReturnIndex should include all the things that comprise the investor return during the holding period including price change, dividends, coupon and bonus issues. The dividends are calculated as being reinvested to the stock. These items were used to calculate the GF rankings and the HPR for the selected stocks in analysis stage 1. In practice the last twelve months cumulative EBIT was used with last interim report balance sheet items and last trading day Enterprise Value from each represented quarter. The number of companies participating in the selection varied with the data available in each quarter, and as new companies were listed and some were delisted.

Delisted companies should be included in the analysis to avoid *survivorship bias* which can occur when only the companies that are still listed when the data is collected are included in the analysis, and those that are for various reasons delisted between that point and the beginning of the study time line are excluded. The purchase prices for the selected companies were taken after the interim information was public. This was ensured by using the TotalReturnIndex value from two months after the quarter end so that most of the time the stock had traded some time after the interim report and before the purchase or sales of the stock from the portfolio. This was done to avoid *look-ahead bias* that occurs when an analysis uses information at time point it would not have been available publicly to investors. The delisted companies are included from 2005 to 2011, but for the older delistings the data was sparse and some deficiencies remained, so in theory there can be some survivorship bias.

However, unprofitable companies do not get selected by GF because EBIT is the nominator in both ranking ratios and negative values remain in the bottom part of the list. Few cases were either Enterprise Value or the invested capital, calculated as sum of Working Capital and PPE as described earlier was negative, were excluded to avoid companies with negative EBIT and some anomalous balance sheet items to get selected. That takes care of bankruptcy delistings. Stocks delisted due to a merger or buyout, on the other hand, probably do not have such inferior returns compared to an average listed company that the expected net influence from the survivorship bias would become anything larger than negligible.

For the second stage analysis, the DM application, a large data matrix with information that was considered relevant for this type of prediction problem was assembled. When the relevance for a variable was weighed, a number of studies with similar tasks were used as a reference, but also many new type of variables were included and derived combining one or more simple data items. With company fundamental data, also Orbis® database was used for some items due to its wider coverage with especially in earlier periods and delisted companies than TF. Macroeconomic data was acquired from Federal Reserve Bank of St. Louis database for economic time series. After enough information had been extracted from the sources, the process continued with the data cleaning which means screening the variables for anomalous values and to get a better understanding of the distribution of the individual variables. This is important in determining their possible value in further analysis as well as requirement for transformations. Many variables that were based on company interim information data were quite non-normal in character, and categorization transformation was found to be the best remedy for this in the sense of the model predicting ability. SPSS Statistics® was used at this phase, and some of the automated tools for the data cleaning task were found useful. Especially the automated categorization tool that optimally sets the cut points for the variable categories. On the other hand the tool that attempts to clean anomalous values by moving far from the center located values routinely closer the statistical mean of the data, were found to worsen the model performance. Likely this is due to loss of information that is in fact valuable, however outlier it may seem, if not examined carefully.

3.4.2 Missing Data

This subsection discusses about common situation where part of the variables in analysis have cases with missing values. Most often this tends to happen with survey data with respondents simply failing or being unwilling to answer some questions. When finding a suitable way of dealing with this issue, it is first important to determine the amount and distribution of the missing data. The suitable action depends e.g. if there are many variables with missing values or some cases that have lot of missing information. But most important it is to find if there is some statistical pattern in the values that are missing. Two basic situations can be distinguished: ignorable and non-ignorable. An example of an ignorable case is when the amount of missing data is less than 5%, and its distribution is random (Tabachnick et al. 2007, 62) i.e. the missing values are not dependent on the other values or completeness of the other observations in the sample. The data is thus said to be missing completely at random (MCAR). This is the simplest case, and all the actions presented in **Figure 9** apply in principle or the cases with missing data can simply be deleted and analyze only the rest of the data (Complete Case Analysis) without introducing any bias into the analysis. Another ignorable case is when the data in the complete cases contains information about the missing values and can be thus used to recover them. This case is called missing at random (MAR). Now some more careful consideration is needed although also deletion of the cases is leads to non-biased sample, like above, as long as the missing data pattern is not conditional on the dependent variable. On the other hand this sacrifices efficiency (Greene 2012, 95).

The missing data is non-ignorable when it is not missing at random (NMAR). For example when surveying expenditure patterns, the high-income class more often fails to report their disposable income. Then the gaps in the data systematically depend on the phenomenon being modeled, and cannot be predicted only from the available variables in the dataset. Then some information necessarily is lost, and no general method for handling this case exists, but more detailed understanding of the process generating the missing data is needed. Selection models have been developed to handle such cases.

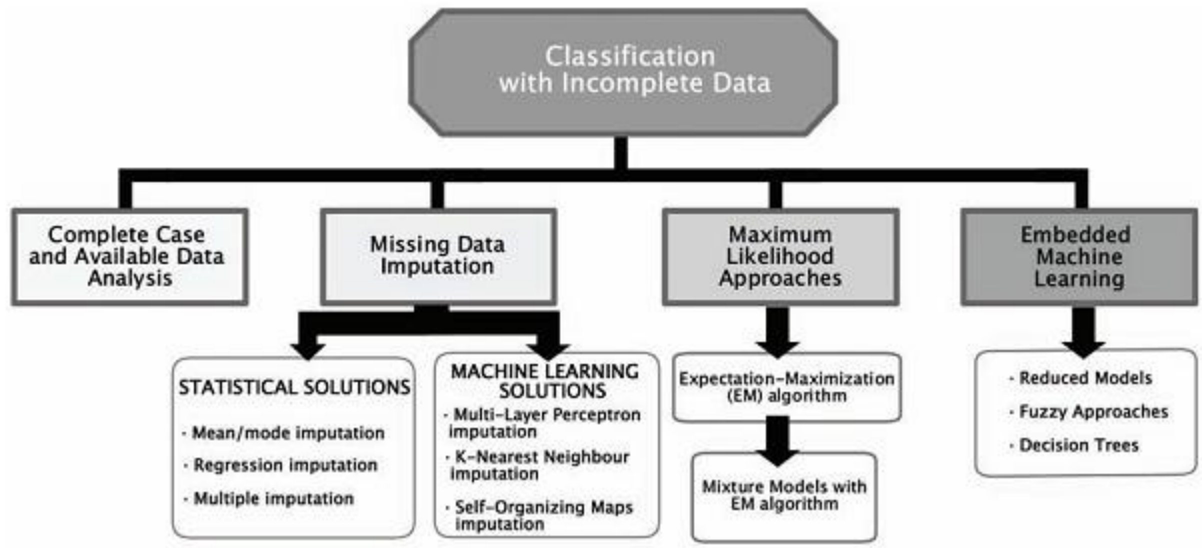


Figure 9: Classification with Incomplete Data (Olivas et al. 2010, 149)

When the missing data is either MAR or MCAR, several possible methods to recover the complete data matrix have been developed. These methods have been presented in **Figure 9**. The first one is to delete listwise the incomplete observations and use only the complete records in the analysis. This is often default choice in statistical software and in case of MCAR it is relatively safe method unless there are few observations. The cases with missing values should be less than 5% of the total data matrix. When deleting is done pairwise, more data is spared as all possible pairs of regressors are used in estimation. Data imputation methods in statistics mean replacing the missing value with some representative or with some statistical methods modeled value. It is then valuable to have a rich data set that contains information to use in predicting the missing values.

Olivas et al. (2010) distinguishes between statistical method imputation and machine learning algorithms. On the other hand growingly popular choice is to use maximum likelihood based methods in modeling the missing values. The last branch in **Figure 9** considers methods that does not actually replace the missing values but attempts the analysis without them. Thus the methods in this group are not suitable for many applications. The statistical solutions are most common, but the simplest mean imputation has obvious disadvantages as it lowers the sample variation and distorts correlations.

Regression imputation – linear or nonlinear – predicts the missing values from the available variables by fitting a line or curve. This method also reduces the variability and adds to multicollinearity. Multiple imputation is the preferred method in this group as it takes into account the uncertainty in the unknown value prediction. A set (e.g. 2-5) of plausible values is generated for each missing value, and the analysis is then conducted normally with each completed data set. The parameter estimates and variances are finally combined for inference. Expectation Maximization (EM) algorithm finds the model parameters that are most likely for the *observed* data, and iteratively finds the maximum likelihood estimates for the missing data.

The amount of missing data in the stage 2 analysis was less than 5%, and it was analyzed with SPSS Statistics® to reveal the possible patterns in it which were not found by visible inspection or descriptive statistics. *Little's* test for MCAR, however, indicated negative for complete randomness. As it was no reason to expect that the missing values would be conditional to HPR (which would have pointed to NMAR data), but on the other hand, variables like Market Value probably could predict the likelihood for missingness, the data was analyzed to be MAR. A deletion of any observations or variables at this point was not considered in an attempt to include all relevant and adequate information to build a good predictive model for HPR of the GF stock selections in stage 1. Also the number of observations was relatively scarce as the adequate ratio of observations to independent variables varies in literature in the range 10-50 depending on the author and modeling algorithm (in the final model of this study it is 230/8 which should then be adequate). Also, not any single variable had over 6-8 % missing. The decision for the imputation method was EM since the data was not MCAR, and the pattern in the missing data was related only to the observed data.

3.4.3 Descriptive Statistics

Total number of the companies that were involved in stage 1 analysis in one or more quarters was 1179, and on average 621 companies participated in ranking every quarter. The GF selected stocks', i.e. the stocks in stage 2 analysis, distribution in market capitalization axis was according to **Figure 10**.

Classification cutpoints are the same than in OMX Nordic exchanges, but it is based on the actual market value in that quarter that the stock was selected and not the official grouping which lags six months behind. The large share of small-cap companies is not surprising since on average it is the small companies that are less analyzed and thus better candidates for undervalued stocks. Large-cap company slice is correspondingly somewhat smaller than in the pooled main lists of the Nordic exchanges.

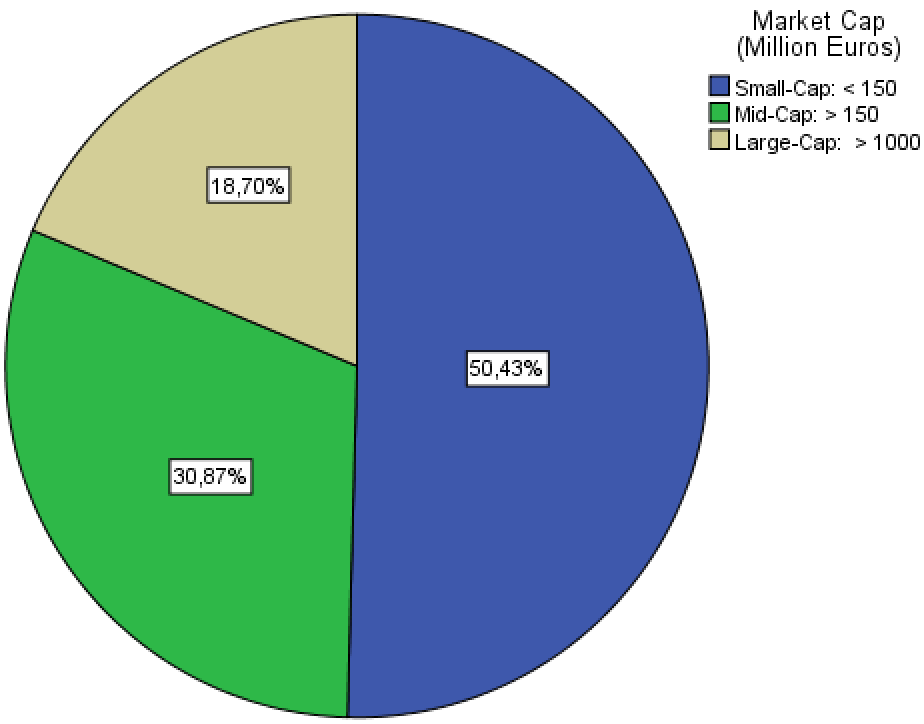


Figure 10: Market Cap Division of the GF Selections

Size dimension is also important to the distribution of the HPR-% variable which is the six month holding period return of the stock selections. This is revealed in **Figure 11** as the shape of the histogram graph is clearly conditional to the Market Capitalization variable.

Most of the extraordinary variation is located in the Small-cap segment, and this should be a good sign from the model building point of view as it should be easier to find the separating variables for Loss and StrongProfit categories when the company size is controlled. Also in the other size segments the very small and large observations seem to be distinctive from the central mass and thus may share some common attributes.

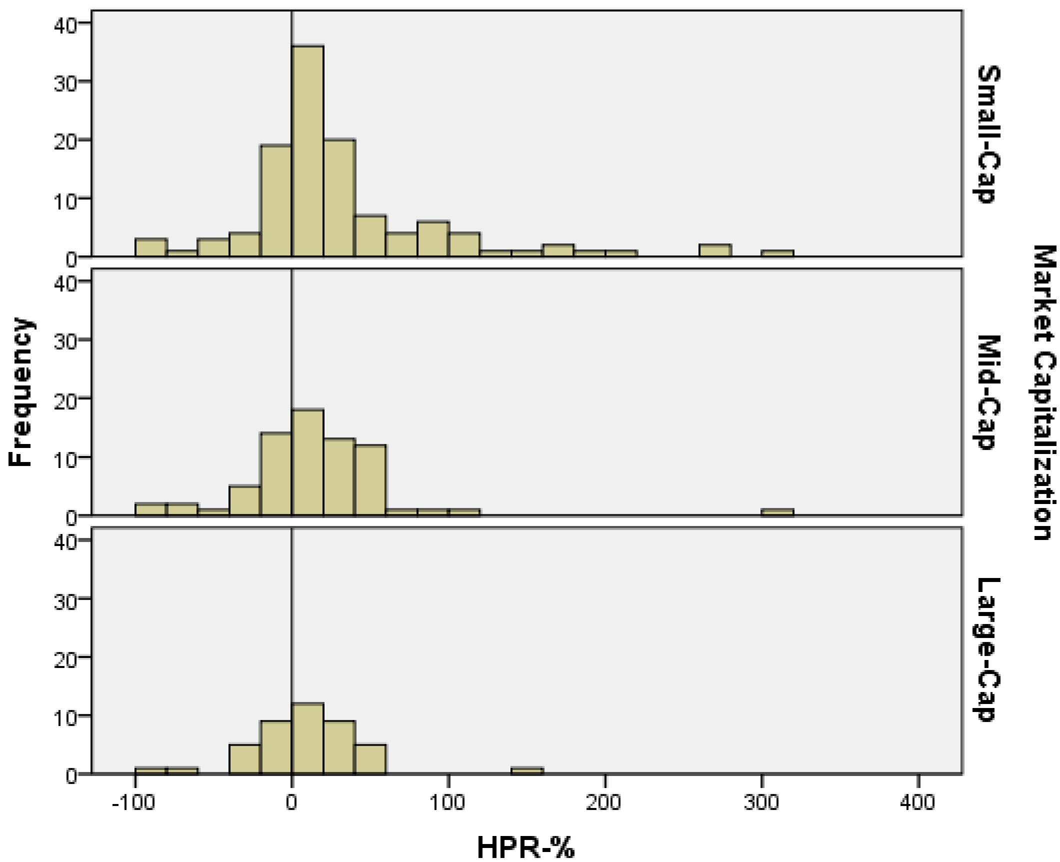


Figure 11: Return-% over Market Cap

Table 3 confirms with numbers the visual observation from **Figure 11:** the Small-cap companies have tripled the expected HPR-% compared to the Large-cap and far wider dispersion which is also positively skewed.

Table 3: Market Cap and Mean Return-%

Market Cap	Mean	N	Min.	Max.	Std. Dev.
Small-Cap	31.4	116	-90.8	315.1	67.1
Mid-Cap	15.6	71	-85.9	300.0	49.7
Large-Cap	10.3	43	-87.4	157.6	38.1
Total	22.6	230	-90.8	315.1	58.0

Figure 12 below show shows the aggregate distribution of HPR-%. As discussed earlier, the distribution is quite hazardous, and in fact, in the following six months from the portfolio formation a significant number of stocks selected with GF reduce in price to a small fraction of their purchase price while some stocks triple in value. This is also Greenblatt's main concern with the practical applications of the formula, thus, for inexperienced investors he suggests diversifying into about 30 stocks and to be confident on that the system works on average. **Table 4** tells this in numbers: although the mean seems to be firmly positive, the standard deviation reaches its triples in magnitude.

Table 4: Descriptives for Return-%

Mean	Median	Std. Dev.	Skew.	Kurt.	Min.	Max.	Percentiles				
							10	30	50	70	90
22.6	13.4	58.0	2.1	7.6	-91	315	-27.2	-.1	13.4	32.4	80.6

Table 4 reveals also that the kurtosis is more than double that of normal distribution and that the distribution is positively skewed. High excess kurtosis means that the probability mass is centered on the mean with heavy tails, and positive skewness that the right tail is very long and thus attractive for investor.

This kind of distribution looks very promising for the task to separate the tales from the center (i.e. separate the winners from losers) with some predictive function. If the distribution was more normal, it would be more difficult to find effective predictor variables since there would not be many attributes that clearly diverges from the central mass. On the other hand with a distribution like this it seems likely that both tails differ significantly in many attributes from the center and from each other. Finding those attributes is the exact task in the second analysis stage. The cutpoints for the classes that are targeted with the separation function, Loss and StrongProfit, are roughly percentiles 30 and 70.

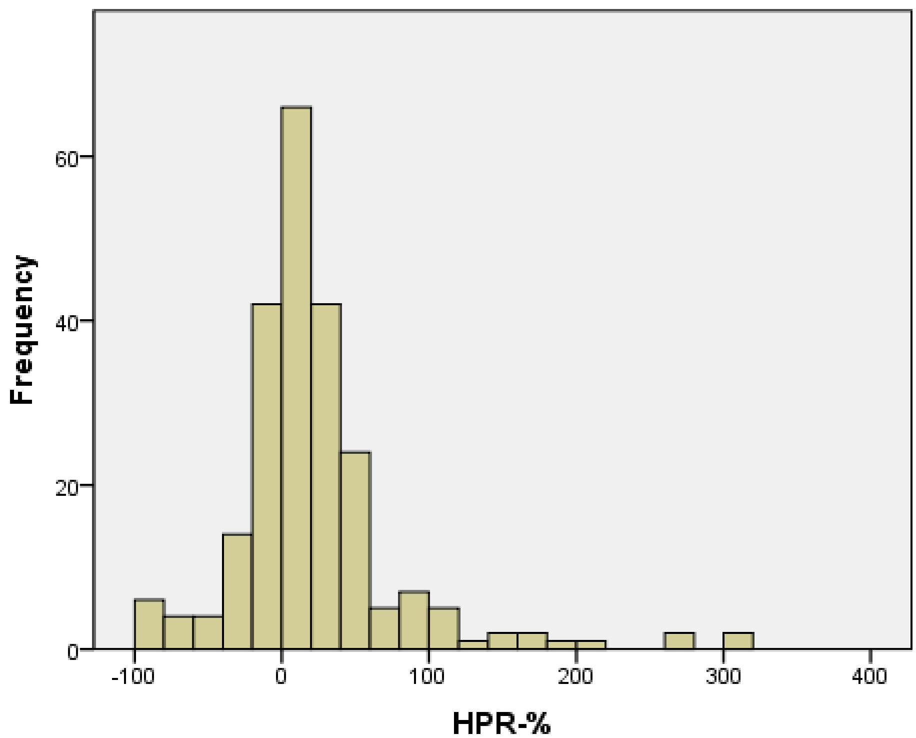


Figure 12: Distribution of Return-%

One very interesting perspective is to find out how these GF selected companies are located in the traditional value stock mapping. **Table 5** provides insight into some key characteristics of the GF selected companies. P/E, P/B and DY are in the range of what would commonly be expected from a value stock but ROE clearly is out of line for most "loser stocks". Same is true for the recent stock price movement prior to purchase; average is plus 15%, and only the lowest quartile is negative.

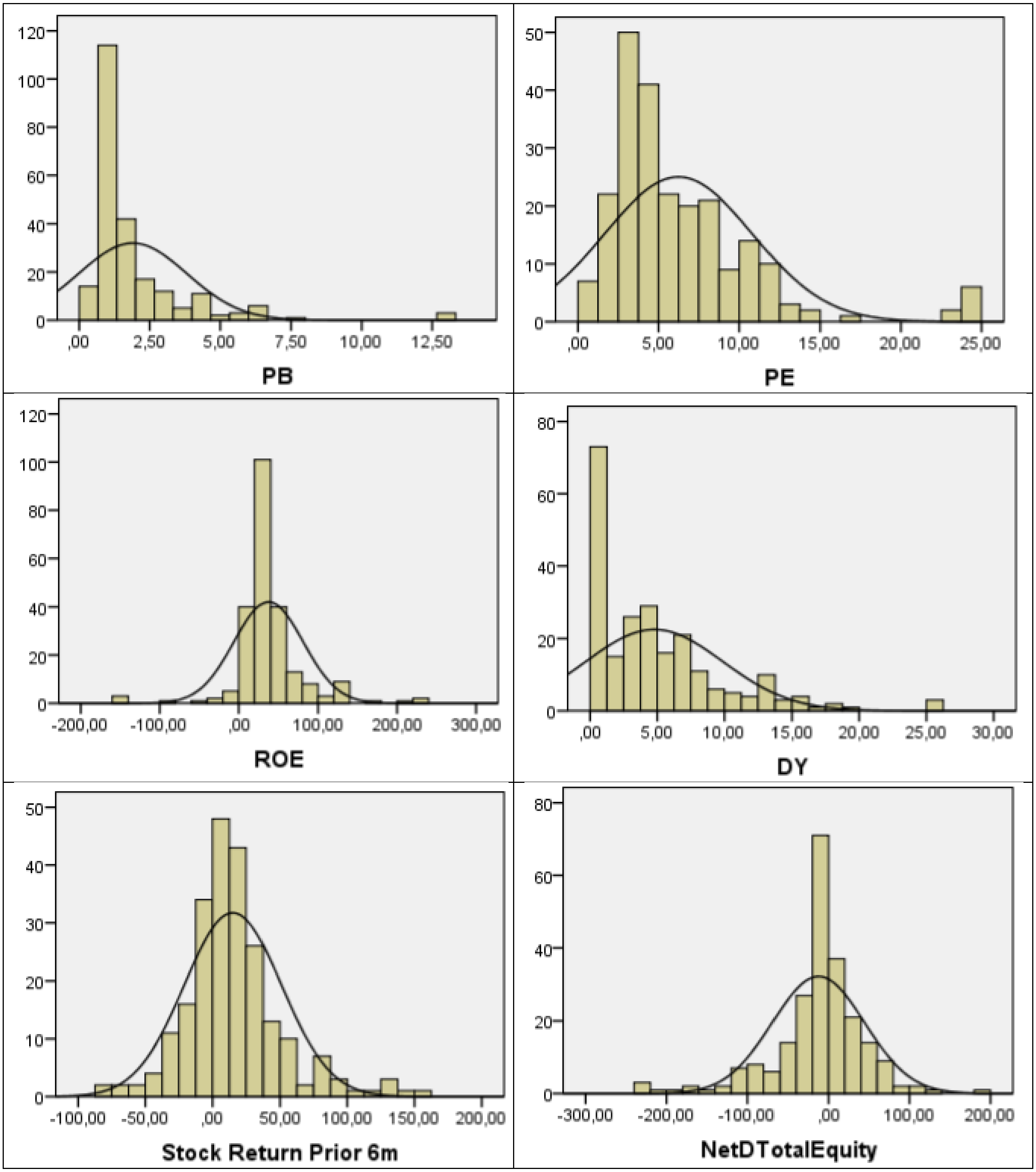
At least part of this movement is probably due to positive earnings surprise. The purchases were executed at a time point two months after the end of quarter, and some earnings momentum was already in effect. Net debtedness also has a promising distribution for a predictor variable candidate. It has a negative mean which tells that a large part of the companies are in healthy financial position while the 75 % quantile is 17 and are thus direly distressed. Intuitive variable selection could start with variables like these when looking for the ones that could separate the future winners and losers in this context.

Table 5: Key Characteristics of GF Selected Companies

Descriptive Statistic	**Mean**	**Standard Deviation**	**Percentiles**		
			25	**50**	**75**
P/E	6.2	4.6	3.0	4.8	8.1
P/B	1.9	1.9	.9	1.2	2.1
ROE %	37.5	43.6	21.2	30.9	48.2
Stock Return % Last 6 Months	15.0	36.1	-3.7	11.4	27.2
DY %	4.8	5.1	.0	3.8	7.0
Net Debt / Total Equity %	-12.5	57.0	-29.6	-5.7	17.0

In **Table 6** the distributions are illustrated with histograms. Visual inspection sometimes is helpful in the search for potential predictor variables, and relying only to automatic tools may not be the best way. Next thing to do would be to test how these variables correlate with the HPR-%. Also new interaction variables can be formed and then test how they perform in a model. This can be done routinely with most modern statistical packages but the variables that are formed this way should be intuitively meaningful. It is thus prudent to constrain the testing only to two-way interactions. From visual and intuitive point of view e.g. the dual combinations of PB, ROE, DY and NetDTotalEquity could be interesting candidates.

Table 6: Distributions Graphs of the Key Characteristics



4. RESULTS

This chapter presents the empirical findings of this study. First section 4.1 presents the findings from analysis stage 1 which is the GF application study in the Nordic Exchanges, and 4.2 covers the results from the analysis stage 2 which is determining a predictive model for the holding period return of the stocks that are selected in stage 1 using DM methods. 4.3 makes inference based on the results and suggests some paths for further research.

4.1 Results from the GF Simulation

4.1.1 GF in Nordic Markets

The program, that was implemented as MS Excel® VBA® application, simulated investments so that first five GF ranked stocks were bought based on interim results at 1Q2000. The price paid was taken two months after the quarter end 1.6.2000 (or nearest trading day). This was done to make sure that the information was available and reflected in the market price at time when the program made its purchase. This was done in every quarter until 4Q2011 when the purchases were terminated. The stocks that were bought remained in the portfolio for six months and were then sold at the market. Trading costs applied were 0.2% per the capital amount of the purchases and sales, and the tax rate for capital net income was set to 29%. Taxes were paid at the end of the following year from the realized gains. Since dividends are included in the TotalReturnIndex (calculated by Thomson-Reuters) which the program used for recording the holding period returns, they were taxed with the same capital income rate than the trading profits. Last stocks were sold and all tax liabilities paid at the market prices of 1.9.2012.

Geometric average yearly rate of return for the studied period was 29.4% p.a. and volatility 39.2%. For the relevant reference index (FTSE Nordic Value Total Return Index) they were respectively 7.6% p.a. and 24.6%. The volatility seems high but there is in practice very little downside movements in the GF portfolio, and standard deviation is not, as was discussed earlier, probably the most suitable risk measure for highly (positively) skewed return distributions. **Figure 13** illustrates graphically the quarterly GF return series alongside with the index return. This index was chosen as a reference index because it is also computed as total return based index, and it consists of Nordic value stocks.

The all-share return indexes have generally performed much worse in this period, and are hardly above the year 2000 levels which was the starting point in this study. There is one very sharp loss in 2008Q3 which of course was common to most markets and instruments due to the global financial crises. The recovery for GF portfolio was also very fast and strong in contrast to the reference index which took a second consecutive dive with same magnitude in the next quarter.

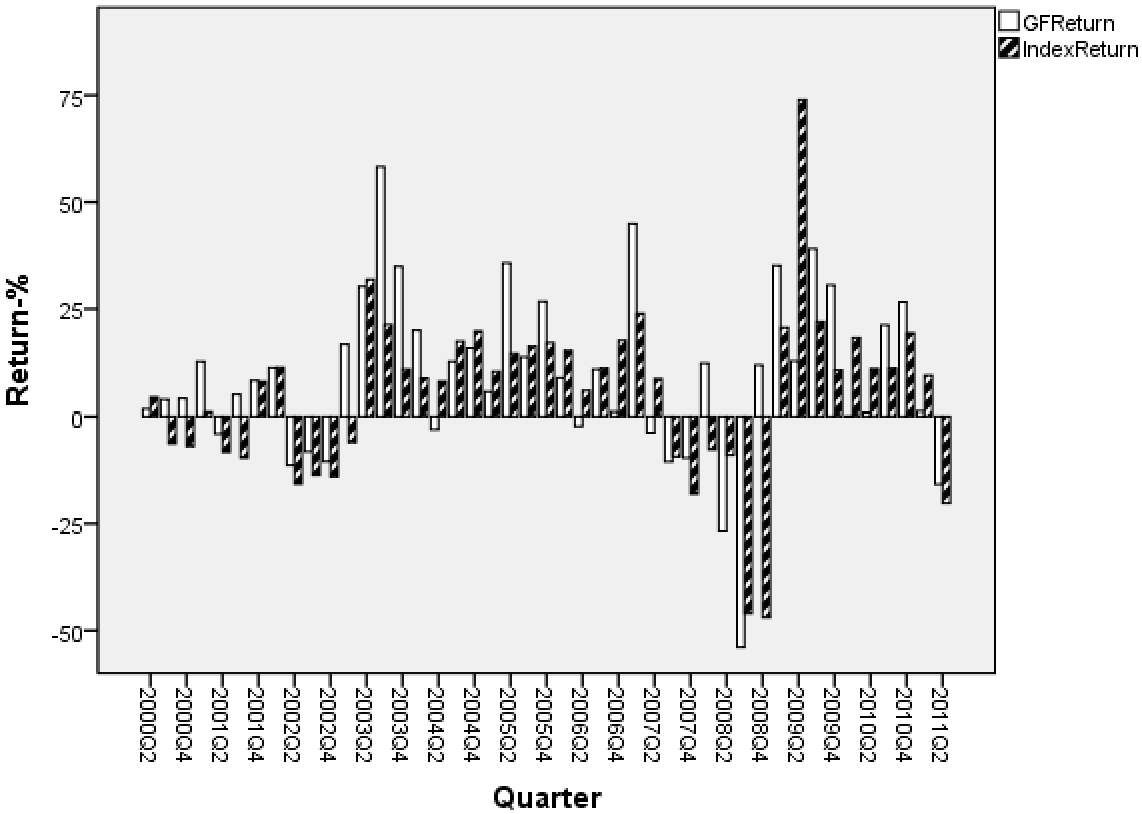


Figure 13: Quarterly GF Returns

When improving GF, first thing is to try with different holding periods and portfolio sizes. Lengthening of the holding period from six months or increasing the amount of stocks purchased at a time from 5 impaired the portfolio returns (this schedule means holding a 10 stock portfolio). Greenblatt, in his book set the holding period to a year because it is intuitive and beneficial for tax optimization purposes under U.S. tax regulations.

The next conventional method would be to experiment with additional value indicators in the ranking and selection criteria. Including other value indicators like P/B did not improve the portfolio return, it was smaller, but it did reduce the volatility. Since most of the volatility anyway is on the upside, losing the return potential was not considered to be acceptable price from illusory safety increase. This result is also in line with the earlier discussion of the papers from Aby (2001) and Elze (2010).

4.1.2 Additional Hypotheses

Some additional hypotheses were also tested. One of them was that the GF selected stocks characteristically resemble value stocks. This has already been established in 3.4.3 Descriptive Statistics not to be exactly true since the sample P/B and ROE means exceeded what is typically observed in value stock group by far. Particularly, those not qualifying the *loser stock* requirement of recently depressed stock price was a surprise as Greenblatt argues that they should be good stocks with cheap price. The stock return in the prior six months to selection was positive on average. This could have been partly due to a positive earnings surprise because the purchases were made two months after quarter end but the stock price development in prior four months was more dominating than in the last two after the end of that quarter when the stock was selected. This suggests that these stocks have already passed the loser stock phase and are gaining some momentum characteristics after a turning point in fundamentals and before becoming overpriced. Also the fact that the HPR-% on average soon deteriorated after six months supports this theory. As a predictor to the HPR-%, the previous six months' stock return was highly significant with a positive coefficient. This demonstrates clearly that there is a momentum effect involved in GF rather than it being a contrarian strategy.

Another interesting hypothesis that was tested was that volatility would increase the GF return, since as discussed, volatility increases the pricing errors necessary for this kind of extraordinary returns. This hypothesis received support at single stock level; the correlation between the prior six month stock volatility and HPR-% was positive and significant although small (Pearson corr. 0.17, Sig. $\approx$0.01).

The market volatility in the previous year also correlated significantly and positively with the GF return in the next year (Pearson corr. 0.43 Sig. < 0.01), as was expected, but alone it was a poor regressor with low coefficient of determination ($\boldsymbol{R^2 \approx 0.2}$) although the constant and the coefficient were both highly significant (Sig. < 0.01). This is graphically observable in **Figure 14** below. On the other hand contemporaneous (same period) market volatility and GF return correlated very strongly and with a negative coefficient. This is not surprising since markets tend to fall when volatility is high and *vice versa*.

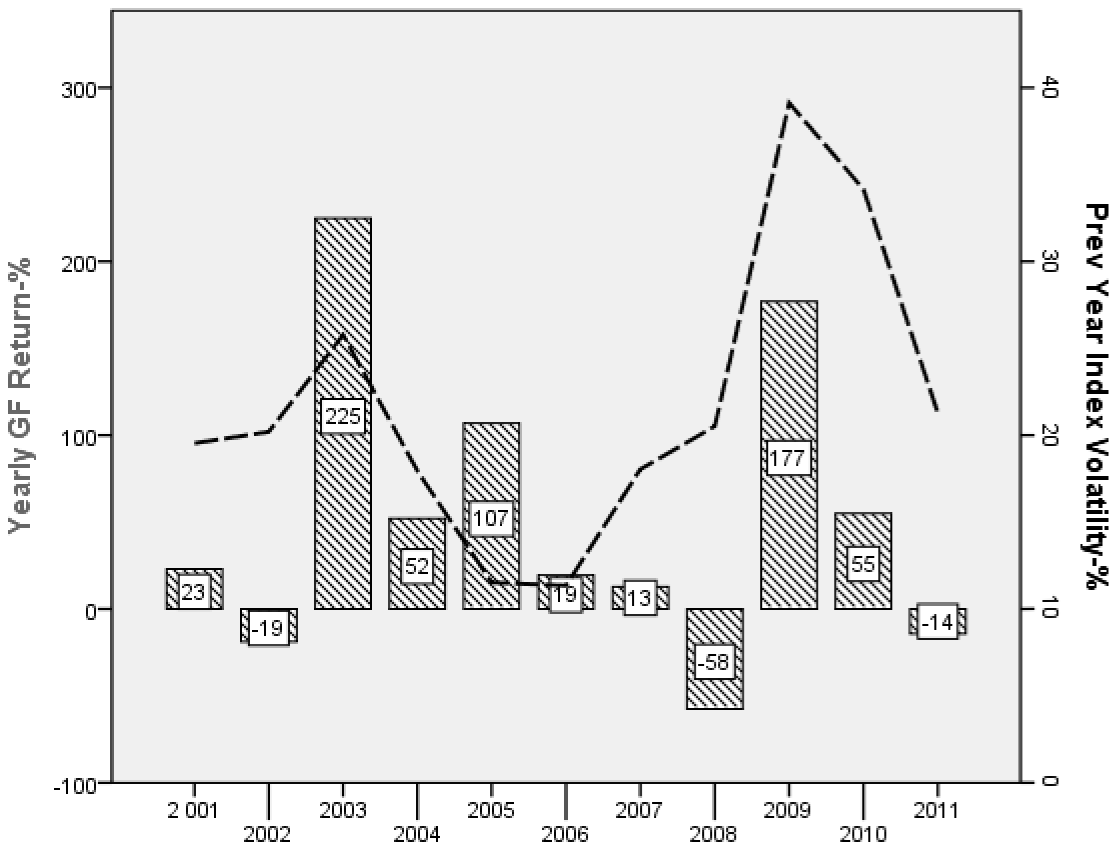


Figure 14: Previous Year Index Volatility and GF Return

4.2 Results from the Predictive Modeling of the Stock Return

4.2.1 Determining the Final Model

In the analysis stage 2, the objective was to find a predictive model by, applying DM methods, that can separate *ex ante* winners from losers in context of the companies that are selected with GF. For this task the holding period return of the companies from the Nordic GF simulation study was coded as 3-class nominal category variable with: 1=StrongProfit (HPR-%≥30, 2=Profit (0≤ HPR-%<30) and 3=Loss. From the initial data matrix a group of influential variables was first extracted. Then a scan for 2-way interactions between these variables was conducted. And those interactions that seemed promising as terms in the predictive model were then examined carefully with and without the main effects (constituent variables) in the model. When the main effects were also significant and included in the model, the interpretations of the interaction term was to indicate variable's varying strength of effect for the response variable conditional on the level of the other constituent variable. If the interaction term was significant but the main effects were not, the term needs to make sense as an independent entity as these terms are no longer interpreted as interactions between variables but new variables. A good example is GNP as a product of price and quantity components in the economy. Nobody thinks that the constituents should be included in the model if GNP is significant predictor. With this data similar thing was with variables like P/E and P/B. They lacked significant influence as independent single variable but were important elements in interaction variables that proxy some more complicated latent company attributes. In the final phase a group of 14 highly influential terms (7 independent variables and 7 interactions) was separated. These terms were then tested with automated tools to determine the final model that would best meet the predetermined criteria: **(1)** Separation performance should be high especially for the Loss- category, **(2)** All model terms, especially the interaction variables, should have intuitive meaning and **(3)** The model should be as parsimonious as possible (include fewer terms). The final model consists of 8 terms and a constant (2 single variables, 6 interaction variables and a constant).

The rate of significant interaction variables to single variables is not surprising; in fact, it was anticipated as the modeling task is far from trivial. The interaction variables represent some underlying factors that can only be measured via a proxy. Screening the 2-way interactions of influential variables and assessing their significance both with statistical methods and intuitively, can discover this new information, and put it to practical use.

The two functions in the model were trained to separate the lower (Loss-category) and upper tail StrongProfit-category) in the HPR-% distribution. The hypothesis was that since the tails seem to be long and distinctive from the center, there would be a distinctive set of attributes that could be *ex ante* predicted with investment time information. The model indeed succeeded very well in this task. The data was divided into three partitions to avoid overtraining, and to ensure that the model performs well with out of sample data. The partitions were training (40%), testing (40%) and validation (20%). With this kind of partitioning the training sample is used for training, testing is used for testing but also refinement of the model and the validation is then used to get the idea how the model performs with out of sample data. The final model was able to predict over all correct 88%, 83% and 87% of the observations in the partitions respectively, and in the validation partition it got correct 90% of the StrongProfit classifications and 95% of the Loss classifications.

4.2.2 Model Fit Statistics

Commonly the first performance indicator to view for a classification model is the overall prediction accuracy table. The correct and wrong predictions of the model are presented by partition in **Table 7**. In this case, as the predictions for the tail categories are weighed more, the best model fit criteria is the matrix that shows the correct and incorrect predictions in each category of the response variable grouped by the partition. This so called *coincidence matrix* can be used for detecting systematic errors in predictions. The coincidence matrix for this model is presented in **Table 8**. In the table the column values are defined by the predicted values, and the rows display the actual observed values. The value is the number of records in each pattern of predicted-actual pair.

The numbers in the overall prediction performance table are already very good but the real value of the model is revealed in **Table 8** as the validation partition classification performance verifies that the model is capable of functioning according its purpose as it can separate the tail observations. If the model predicts that the stock will be in StrongProfit-category, it is correct 90% of the time. Further, when the stock is predicted to be in Loss-category, the model accuracy is nearly 95%. On the other hand if the prediction is in the center category, Profit-category, the correct categorization percent drops to 78% in validation participation.

An indicator of the model consistency is also that there are no cross-tail classification errors in the validation partition i.e. stocks that actually belonged to Loss-category would have been predicted to be StrongProfit or vice *versa*. Only in the training and testing partitions is one such misclassification but none in the final phase. The consequences for even one notch misclassifications in applications for Loss-category could be more severe than for StrongProfit-category since in aggressive strategy they would be sold short. On the other hand, a long investment in a stock predicted to be a high-rising and realizing only moderately positive (0-30%), is not a disaster.

Table 7: Overall Prediction Performance

Partition	**Training**		**Testing**		**Validation**	
Correct	77	87.5%	66	82.5%	54	87.1%
Wrong	11	12.5%	14	17.5%	8	12.9%
Total	88		80		62	

TABLE 8: The Coincidence Matrix

Partition = Training	**StrongProfit**	**Profit**	**Loss**
StrongProfit	14	0	0
Profit	2	42	3
Loss	1	5	21
Partition = Testing			
StrongProfit	16	2	1
Profit	3	34	1
Loss	0	7	16
Partition = Validation			
StrongProfit	15	4	0
Profit	2	18	1
Loss	0	1	21

The observation that the model performs better in discriminating the loss producing stocks is intuitive since, as a group, the stocks that rank high in GF selection and nevertheless end up in negative result understandably share a set of distinctive characteristics. The cut-point between the two positive categories, on the other hand, was more arbitrary thus higher prediction error is expected. There are more errors in predicting Profit-category stocks. This is not surprising as it was the stocks in the return distribution tails that were supposed to possess the distinctive characteristics. Neither is it alarming since in the practical applications of the model, the investments would target on the tail categories with positive or negative weights. It is also consistent that the model improves in tail category prediction accuracy in each analysis step from training to validation as there is some parameter adjustments conducted even in the testing partition when validation is used. Although in the center category accuracy there is a slight deterioration from the training to the testing and validation, the tail accuracy improves further which is the main objective that the model was designed to accomplish. It can thus be misleading to target primarily to the overall model accuracy as the model performance bench mark.

SPSS Modeler provides also statistical model fitting information. In **Table 9** is presented some standard relative indicators of the whole model significance. The likelihood ratio provides a test of the final model against the model with constant only. The smaller the number in the Final-row the better as it represents less uncertainty in the full model. Null-hypothesis is that the coefficients of all the other terms than the constant are zero, and Chi-Square-column shows how much evidence the data provides against that hypothesis. The figure in the Sig.-column is the risk level for the rejection of the null (p-value). So the model clearly is significant. AIC (Akaike Information Criterion) is an information theory based model fit indicator. The final model has smaller value which indicates better fit than in the intercept only model. Pearson and Deviance values work in opposite direction, and the lack of significance indicates better model fit. The three metrics under Pseudo R-Square –heading are developed to be similar with the $\boldsymbol{R^2}$ – coefficient in linear regression, and bigger number indicates a better fit. They do not, however, measure the ratio of unexplained variance to the one explained by the model, and should thus be interpreted with caution.

Table 9: Goodness of Model Fit

Model	**Criteria**		**Likelihood Ratio Tests**	
	AIC	**-2 Log Likelihood**	**Chi-Square**	**Sig.**
Intercept Only	490.1	486.1		
Final	363.8	159.8	326.3	.000
Pearson			190.8	1.000
Deviance			159.8	1.000
Pseudo R-Square				
Cox and Snell: 0.76		Nagelkerke: 0.86	McFadden: 0.67	

To monitor the significance of the individual model terms, SPSS offers likelihood ratio tests. Similarly to the whole model case, the test measures the difference between -2 log-likelihoods of the model that includes the parameter in question and the one where it is omitted (reduced model). Chi-square statistic measures the effect of the parameter and Sig. - column shows the significance p-value for the contribution of the parameter. Like usually the p-value smaller than 0.05 means that the model fit is impaired if the term is omitted. In **Table 10** are the model terms and their likelihood ratio tests. Bigger chi-square value means stronger effect of the term in the model and more evidence against the null-hypothesis that all the term parameters are zero. The observation data for the variables included in the model is displayed in **APPENDIX 1**.

As the **Table 10** shows, all the model terms excluding the constant are highly significant. The dot in the Sig.-column for the constant means that the intercept term could be omitted from the model without deteriorating its performance. However, the constant should be included in multinomial logistic regression as it represents the log-odds between the response categories and the base category when all the function coefficients are zero. This value deviates from zero unless the groups are exactly similar in size so the regression line should not be forced to through the Origo. The categorical variables (N or O coded) are represented in the model by $\boldsymbol{k-1}$ dummy variables for one less than the amount of categories$\boldsymbol{k}$ of the variable. Interactions between the interval (I coded) and the 1/0- valued dummy variables are then simply either zero or the value of the interval variable. Similarly the interactions between the categorical variables are the cross-products of the sets of 1/0 dummy variables.

Table 10: Likelihood Ratio Tests

Model Term	Chi-Square	Sig.
Constant	0	.
A21O	25.1	.000
C5I	41.2	.000
A13O*A29O	84.41	.000
A17O*A2O	70.5	.000
A35O*A34O	68.0	.000
C1N*A35O	97.8	.000
C2N*A4O	126.8	.000
A8O*B12I	39.7	.000

When the number of explanatory variables is increased, it often improves the model fit in training partition because of overfitting. This happens when the model parameters adapt to the random noise in the dependent variable. This usually also impairs the model prediction capability in out-of-sample data like validation partition. Another symptom of an over defined model is also that the independent variables often become correlated. This Multicollinearity has adverse effects to stability and reliability of logistic regression models in particularly hence it is common to test for its presence. The most common method of testing for multicollinearity is pairwise correlation between the independent variables but it is not always sufficient since multiple variables can have linear dependencies even if it is not revealed by the pairwise correlogram. With collinearity sensitive methods it is thus customary to calculate VIF indicator (variance inflation factor) or its inverse, tolerance, which tells the proportion of variance in the explanatory variable that is not explained by the other independent variables ($1-R^2$). VIF is the factor by which the coefficient estimators' variance increases. Tufféry (2011, 88) states that 0.2 is commonly considered to be minimum acceptable for tolerance or at least it should be above 0.1 ($VIF \leq 10$).

In this model there are a couple of dummy variables that report above limit VIF-scores but none of the model terms seems to be too seriously affected by this as the correlated parts represent only a small fraction of them. Anyway the model performs well also with the out-of-sample data and thus the main concern of multicollinearity in the model is not a problem.

4.2.3 Enhanced Greenblatt's Formula

After the final model was determined, SPSS Modeler's deployment function scored all the observations based on the model's prediction. **Figure 15** shows how the filtration works with the data. The HPR-% distribution is now paneled by the predicted classification, and it is plain to see that the model has worked as anticipated. The distribution of the observations that were *ex ante* classified as StrongProfit is quite neatly positioned along the positive x-axis towards the high-return end. The mid-section is also quite well handled but there is some more leakage to both sides, and it includes even two quite heavy value losers. Finally the left-tail is very well captured by the Loss-categorization with only minor leaking to the Profit-domain. This of course is problematic if the Loss-categorized stocks are sold short in practical application but it can be handled with *stop-loss* strategies. If these stocks are just ignored and left out from the GF investment portfolio, it should alone induce a considerable return boost. The good thing is that there are no cross-tail classification errors on either way. To further clarify the model performance, **Table 11** displays in numbers the HPR-% distribution conditional on the Predicted Score.

Table 11: Mean Return-% Conditional on Model Prediction

Predicted Score	Mean HPR-%	N	Std. Deviation	Minimum HPR-%	Maximum HPR-%
1	92.4	53	73.8	-8.2	315.1
2	14.3	113	20.1	-83.4	52.3
3	-20.7	64	28.7	-90.8	51.9
Total					

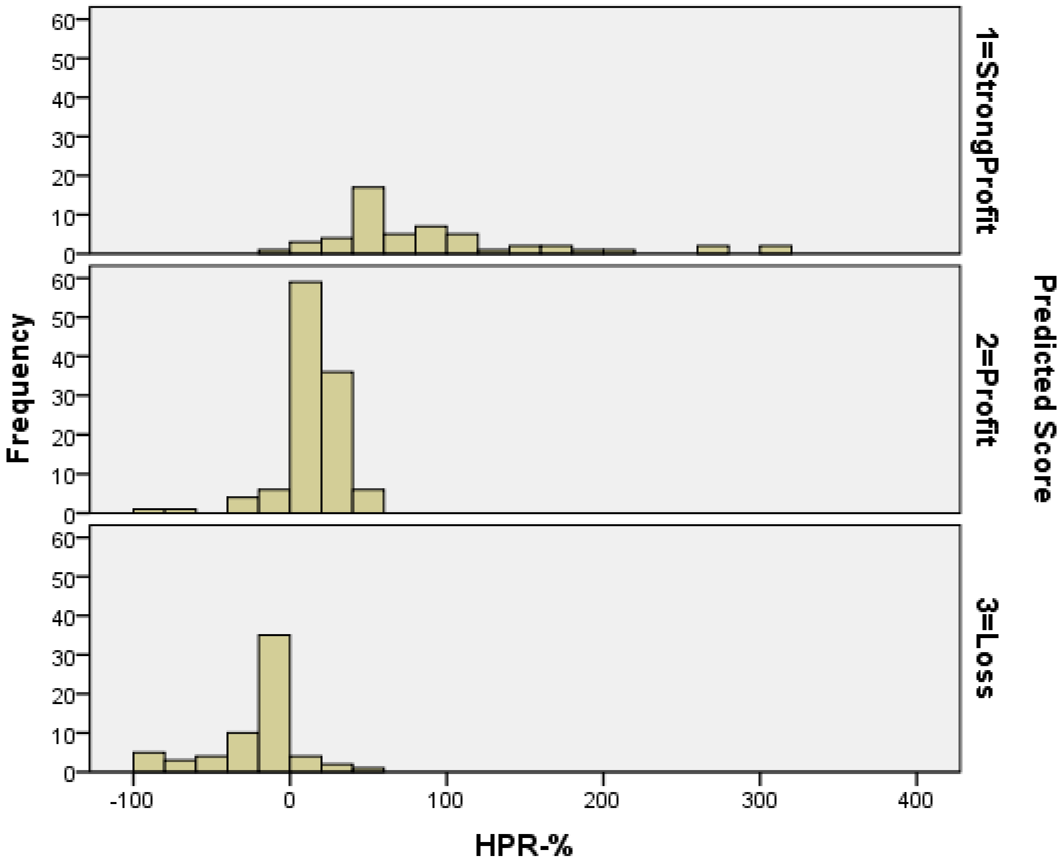


Figure 15: Model Prediction and Return-% Distribution

To test how the predictive model would effect to the GF portfolio return is used as filter in the stock selection phase, the simulation program was adjusted to use the prediction scores and **(1)** simply ignore the Loss categorized (category-3) stocks as short selling was not in practice possible for the majority of the stocks at the period studied **(2)** to increase the portfolio weight for the stocks pre-classified as Profit or StrongProfit.(category-2 and category-1 respectively). It would have been convenient to invest only in the category-1 stocks but as seen from **Table 11**, they are relatively scarce, and were not available in every quarter in this sample. Also the economic conditions affect the occurance of the category-1 stocks as the model term C5I is the first principal component of a matrix with certain economy based information from the Nordic countries, and it has a strong effect in the predictive function for the category-1 stocks. This is the reason why this model does not find these stocks in every quarter, and then many in some other quarters, when the economy has changed.

This economic factor is not significant in the function for the category-3 stocks but has an opposite sign as intuitition suggests. The attributes that best distinct the category 3-stocks are more dominantly company based and idiosyncratic than dpending on market conditions. In real-life application the scoring would be extended beyond the five most highly ranked stocks but with this test they were the only ones in the investment universe. To be able to conduct the simulation by the same principles as with the raw GF experiment and stay fully invested in every quarter, also the category-2 stocks were allowed to enter the portfolio. After all they have a positive expected return, as can be seen in **Table 11**. In every quarter the category-3 stocks were eliminated, and if category-1 stocks existed, their weight was increased until all the available funds were invested. Were they not present, the category-2 stocks were used the same way.

The results were impressive to say the least. For the whole research period the geometric annual average return was 43.8% and volatility 39.8%. The volatility seems high but this time the *de facto* down side risk has been significantly reduced relative to the raw GF invesment portfolio or the reference index. **Figure 16** presents the yearly return bars from the enhanced GF simulation experiment. There are are as many unprofitable years, and the highest returns are not higher than with the raw GF yearly returns, but the lows are not as deep (**Figure 14**, p.87). The average annual return rises from 29.4% to 43.8% in the period 2000-2011 when this predictive model is used to enhance the basic GF based investment portfolio.

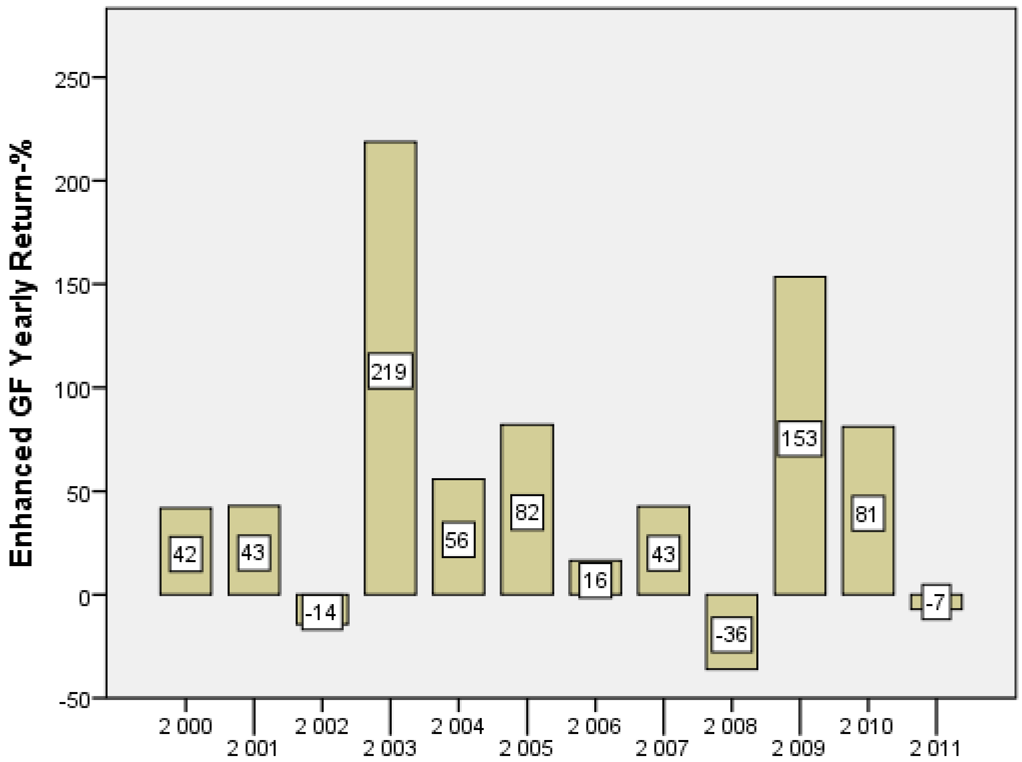


Figure 16: Enhanced Greenblatt's Formula Annual Return

4.3 Discussion of the Results

4.3.1 Summary of the Main Results

The literature review chapter discussed about the differences in thinking of the market efficiency – behavioral schools of thought and the consequences for practical work in finance when one or the other of the approaches is accepted. The view taken in this text is from the side that accepts the inefficient markets thus not all the opposite side's arguments were thoroughly considered. The purpose of this discussion was to lay ground for the empirical part of the text by reviewing a few notable papers with results that lead to the conclusion that with certain company specific information set it is possible to select into portfolio those stocks that return extraordinarily well without increasing the *de facto* riskiness of the portfolio.

This view is very different from the conventional finance where beating the index return consistently is not possible. However, with the evidence presented in the chapter, the conclusion was that this view may be far too narrow, and some new theories about origin and characteristics of returns must emerge.

In the empirical part of this text the objectives were **(1)** to determine whether the Joel Greenblatt's stock selection method that has worked with considerable success in the U.S. markets would produce comparable results in the Nordic markets **(2)** use DM methods to enhance the return distribution of the stocks selected by the Greenblatt's Formula (GF) by predictive classification. This study shows that a simple GF application in years 2000-2011 was able to produce an annual return of 29.4%. A passive value stock strategy, represented by FTSE Nordic Value Index, returned 7.6% p.a. in the same period. The respective volatilities as standard deviations of the quarter returns were 39.2% and 24.6% annually. The volatility of the GF portfolio appears high but, as argued earlier, standard deviation is problematic as a measure of portfolio riskiness when the returns are not normally distributed. The GF return distribution is right skewed so that the volatility is dominantly on the upside. The GF performance in the Nordic stock exchanges is impressive and parallel to the results Greenblatt had when he tested the formula in the United States' exchanges 1988-2004. This suggests, for one, that the results are not coincidental as the studied regions are totally separate, and the research time frames are only for a small part over lapping. Greenblatt himself explains the success of his formula with behavioral phenomena that affect mostly the professional market participants in money management industry, and affect in a way that the fund managers are prone to ignore a certain types of stocks. As the Nordic markets are smaller and more volatile, more inefficient, than those in the U.S., GF portfolio should perform much better compared to the market index in there. This would be in line with the Leivo and Pätäri (2011) paper where they also come to that conclusion with their own enhanced value portfolio. The results in this study, in fact, agree to this as in Greenblatt's study the GF portfolio return was 30.8% p.a. and SP500 12.4% p.a. The return difference with the reference index is far narrower than in this study concerning the Nordic markets.

The second objective was to further enhance GF with a predictive classification filtration. This aim was achieved by applying DM methods to a data set that would have been available at the investment time. The logistic regression model with constant and eight additional terms, that was defined, was able classify the stocks selected by GF to value losers with 95% accuracy and to very high returning stocks (HPR≥30%) with 90% accuracy. When the simulation was recreated using this model in the stock selection phase, the annual return for the research period rose to 43.8% and the volatility to 39.8%. The volatility, though it may seem high, was now strictly concentrated on the upside. The simulation setting was limited as in real production application more data would be available, and the scoring would be extended to wider stock universe as explained above in more detail. The weighting scheme used was also arbitrary but logical. There is, however, no reason to expect the results in production environment to be in any relation inferior. The results in this second stage analysis are very much in line with those earlier papers in the field of contextual fundamental analysis discussed earlier that show that it is easier to find powerful return predictors for stocks when they are first selected or sorted in some meaningful way. As shown in this text, DM techniques provide relevant tools for information extraction in the phase where variable set and functional form for the predictive model are determined.

4.3.2 Suggestions for Further Study

Some interesting questions are raised but left unanswered at this point. Most important would be to find a new comprehensive asset pricing theory to replace the old models of the Modern Financial Theory that are too often based on set of simplifying assumptions and thus are not usable in practical work. The behavioral finance, on the other hand, has a set of rules and theories about human behavior but lacks structure that could be applied to quantitative asset pricing exercises. One possible approach is the behavioral SDF that was introduced earlier as a synthesis between MFT and behavioral schools of thought. It would offer more structured way to introduce irrationality into asset pricing. The problem is that the new models are not likely to be as “neat” and easily presentable as those of CAPM for example. One target for further research is the Greenblatt’s Formula itself. What is it in the companies’ real processes and conditions that is common among those stocks that the formula selects, and what is the role of the differences in the financial statement practices?

5. CONCLUSIONS

I have in this text showed that investors have choices to their portfolio strategies beyond passive indexing, and that they can be rewarded for stock picking even with rule-based systems that are simple to implement and require very little effort and sophistication. This of course violates some previously powerful theories of the bygone era that leaves investors with a choice of diversification of his funds between interest and stock index funds. This text has reviewed many more contemporary alternatives for such overly simplistic frameworks and showed that there is well-founded motivation to consider more realistic yet unfortunately more complicated distribution, risk and preference models for evolving future asset pricing theories. The old theories are still taught surprisingly widely notwithstanding the mounting conflicting evidence and even abandonment by their original fathers. I have done my best to discredit these tenets **(1)** by reviewing studies that show that the underlying assumptions in these theories are invalid; **(2)** reviewing theories of alternative schools of thought that provide the way around of the deficiencies of the traditional models and fit better to empirical data; **(3)** presenting studies that show how investing with anomaly utilizing or even ad-hoc methods that use contextual and data driven approaches can generate returns that far exceed those achievable if markets were efficient in the sense that the prevailing consensus of the late last century stated. This study has its roots in enhanced value strategies and contextual fundamental analysis supplemented with data mining techniques that are rapidly making their way into the frontline of applied and empirical finance and econometrics. *"Of course, given the quantity of historical data that are now available, optimal forecasts of stock returns going forward may place greater weight on the data, and less weight on theoretical restrictions, than those methods that most successfully predicted stock returns during the twentieth century"* Campbell (2008).

REFERENCES

Agresti Alan 2007. *An Introduction to Categorical Data Analysis*, Wiley

Andrei Shleifer, Robert W. Vishny, 1997, The Limits of Arbitrage, *The Journal of Finance, Vol. 52, No. 1,* 35 55.

Ang Andrew, Hodrick Robert J., Xing Yuhang, Zhang Xiaoyan, 2006, The Cross-Section of Volatility and Expected Returns, *The Journal of Finance, Vol. LXI, No. 1*, 259-299.

Artzner, P., Delbaen F., Eber J., Heath D., 1999, Coherent Measures of Risk, Mathematical Finance 9, 3, 203.

Asness, Cliff S., Tobias J Moskowitz, Pedersen Lasse H. *"Value and Momentum Everywhere"*, February 2009. http://pages.stern.nyu.edu/~lpederse/papers/ValMomEverywhere.pdf

Balvers Ronald J., Wu Yangru, 2006,: Momentum and Mean Reversion Across National Equity Markets, *Journal of Empirical Finance 13*, 24-48.

Banz Rolf W., 1981,The Relationship Between Return and Market Value of Common Stocks, *Journal of Financial Economics, Vol. 9 Issue 1*, p3-18.

Barberis Nicholas, Huang Ming, 2008, The Implications of Probability Weighting for Security Prices, *American Economic Review 98,* 2066-2100.

Barberis Nicholas, Thaler Richard H., 2002, *A Survey of Behavioral Finance*, Available at SSRN: http://ssrn.com/abstract=327880 or http://dx.doi.org/10.2139/ssrn.327880.

Bartov E., Kim, M., 2004, Risk, Mispricing, and Value Investing, *Review of Quantitative Finance and Accounting, 23*, 353-376.

Basu S., 1977, Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A test of the Efficient Market Hypothesis, *the Journal of finance, vol 32, no 3*, 663-682.

Benartzi Shlomo, Thaler Richard H., 1995, Myopic loss aversion and the equity premium puzzle, *Quarterly Journal of Economics, Vol. 110 Issue 1,* 73-93.

Benos, Alexandros V., 1998, Aggressiveness and Survival of Overconfident Traders, *Journal of Financial Markets, 1(3– 4)*, 353– 83.

Biglova A., Ortobelli S., Rachev, S.T., Stoyanov, S., 2004, Different Approaches to Risk Estimation in Portfolio Theory. *Journal of Portfolio Management 31*, 103-112.

Bird R, Casavechia L. 2007, Sentiment and Financial Health Indicators for Value and Growth Stocks: the European Experience, *European Journal of Finance; vol.13*, 769-793.

Bird R, Whitaker J, 2004, The Performance of Value and Momentum Investment Portfolios: Recent Experience in the Major European Markets: Part 2. *Journal of Asset Management; 5,* 157-175.

Bradford De Long J., Shleifer Andrei, Summers Lawrence H., Waldmann Robert J., 1998, Noise Trader Risk in Financial Markets, *The Journal of Political Economy, Vol. 98, No. 4,* 703-738.

Bradshaw Mark T, 2004, How Do Analysts Use Their Earnings Forecasts in Generating Stock Recommendations?, *The Accounting Review Vol.79 No. 1*, 25-50.

Campbell John Y., 2000, Asset Pricing at the Millennium, *Journal of Finance 55, no. 4:* 1515-1567.

Campbell John Y., Hilscher Jens, Szilagyi Jan, 2008, In Search of Distress Risk, *Journal of Finance Vol. 63, 6*, 2899-2939.

Campbell John Y., Kyle Albert S., 1993, Smart Money, Noise Trading and Stock Price Behaviour, *Review of Economic Studies, Vol. 60 Issue 1*, p1.

Campos Julia, Ericsson Neil R., Hendry David F, 2005, General-to-specific Modeling: An Overview and Selected Bibliography, *Board of Governors of the Federal Reserve System International Finance Discussion Papers Number 838*.

Chan Louis K.C., Lakonishok. Josef, 2004, Value and Growth Investing: Review and Update, *Financial Analysts Journal, 60:1*, 71-86.

Clarke Bertrand, Fokoué Ernest, Zhang Hao Helen 2009. *Principles and Theory for Data Mining and Machine Learning*, Springer

Clubb C., Naffi M., 2007, The Usefulness of Book-to-Market and ROE Expectations for Explaining UK Stock Returns, *Journal of Business Finance & Accounting, 34(1) & (2)*, 1–32.

Cochrane John H., 2005, Financial Markets and Real Economy, *Foundations & Trends in Finance, Vol. 1 Issue 1*, 1-101.

Cochrane, John H., 1991, Production-Based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations, Journal of Finance, Vol. 46 Issue 1, 209-237.

Coggin T. D., Fabozzi F. J., 2003, 3 rd. ed., *Handbook of Equity Style Management*, John Wiley & Sons Inc.

Cox D. R, Donnelly Christl A. 2011, *Principles of Applied Statistics*, Cambridge University Press

Cuthbertson Keith, Nitzsche Dirk, 2004, *Quantitative Financial Economics*, John Wiley and Sons.

De Vries Casper G., Danielsson Jon, Jorgensen Björn N. , Mandira Sarma 2006, Comparing Downside Risk Measures for Heavy Tailed Distributions, *Economic Letters Vol.92 Issue 2*, p.202-208.

DeBondt Werner F. M., Thaler Richard, 1985, Does the Stock Market Overreact?, *Journal of Finance, 40, 3*, 193-805.

Dichev Ilia D., 1998, Is the Risk of Bankruptcy a Systematic Risk?, *Journal of Finance. Jun98, Vol. 53 Issue 3*, p1131-1147.

Dowd K., Blake D., 2006, “Measuring market risk”, John Wiley & Sons.

Dutta A., Bandopadhyay G., Sengupta S, 2012, Prediction of Stock Performance in the Indian Stock Market Using Logistic Regression, *International Journal of Business and Information Volume 7, Number 1*, 105-136.

Easton Peter D., 2004, PE Ratios, PEG Ratios, and Estimating the Implied Expected Rate of Return on Equity Capital, *Accounting Review, Vol. 79 Issue 1*, 73-95.

Elze Gregor, 2010, Value investing anomalies in the European stock market: Multiple Value, Consistent Earner, and Recognized Value, *Quarterly Review of Economics & Finance, Vol. 50 Issue 4*, 527-537.

Fama Eugene F., French Kenneth R., 1992,The Cross-Section of Expected Stock Returns, *Journal of Finance, Vol. 47 Issue 2*, p427-465.

Fama Eugene F., French Kenneth R., 2004, The Capital Asset Pricing Model: Theory and Evidence, *Journal of Economic Perspectives, 18, 3*, 25-46.

Fama Eugene F., French Kenneth R., 2006, The Value Premium and the CAPM, *Journal of Finance, 61, 5,* 2163-2185.

Fama Eugene F., MacBeth James D.,1973, Risk, Return, and Equilibrium: Empirical Tests, *Journal of Political Economy, Vol. 81 Issue 3,*607-636.

Fama, E. F., 1970, Efficient capital markets: A review of theory and empirical work, *Journal of Finance, 25*, 383–417.
Fama, E. F., Roll, R.,1968, Some Properties of Symmetric Stable Distributions, *Journal of American Statis. Assoc. 48*, 817-36.

Fama, Eugene F., French, Kenneth R., 1988, Permanent and Temporary Components of Stock Prices, *Journal of Political Economy, Vol. 96 Issue 2*, p246.

Fox Justin 2011, *The Myth of the Rational Market – A history of Risk, Reward and Delusion on Wall Street*, Harper- Collins Publishers.

Francis J, Schipper K., Vincent L., 2003, The Relative and Incremental Explanatory Power of Earnings and Alternative (to Earnings) Performance Measures for Returns, *Contemporary Accounting Research*, *Vol. 20, Issue 1*,121-164.

Francis Jack C., Kim Dongcheol, 2013, Modern Portfolio Theory + WS: Foundations, Analysis, and New Developments, John Wiley & Sons, Inc.

Fuertes A., Coakley, J., 2004, Momentum and Reversal Dynamics in US Valuation Ratios, *Essex Finance Center Discussion Paper 0413.*

Ganzach Y., 2000, Judging Risk and Return of Financial Assets, *Organizational Behavior and Human Decision Processes, 83*, 353–370.

Gilli M., Kellezi E., Hysi H., 2006, A Data-Driven Optimization Heuristic for Downside Risk Minimization, *Journal of Risk, Vol. 8 Issue 3*, 1-18.

Giudici Paolo, Figini Silvia 2009. *Applied Data Mining for Business and Industry*, Wiley

Gode D., Mohanram P., 2003, Inferring the Cost of Capital Using the Ohlson–Juettner Model, *Review of Accounting Studies, Vol. 8 Issue 4*, 399-431.

Goovaerts M.J., Kaas R., Dhaene J., Tang Q., A Unified Approach to Generate Risk Measures, *ASTIN Bulletin, 33(2)*, 173–192.

Gromb Denis, Vayanos Dimitri, 2010, Limits of Arbitrage: The State of the Theory, *INSEAD Working Papers Collection, Issue 23*, preceding 1-32.

Guay W., 2000, Discussion of Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers, *Journal of Accounting Research, Vol. 38 Issue 3*, 43-51.

Gul Faruk, 1991, A Theory of Disappointment Aversion, *Econometrica,Vol. 59 Issue 3*, p667-687.

Hirshleifer David, 2001, Investor Psychology and Asset Pricing, *Journal of Finance, Vol. 56 Issue 4*, 1533-1597.

Jackwerth J. C., Rubinstein M., 1996, Recovering Probability Distributions from Option Prices, *Journal of Finance, 51*, 1611–1631.

Jegadeesh Narasimhan, Titman Sheridan, 1993, Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance, Vol. 48 Issue 1*, p65-91.

Jensen Michael C., Black Fischer Scholes Myron S., 1972, The Capital Asset Pricing Model: Some Empirical Tests, Studies in the Theory of Capital Markets, Praeger Publishers Inc.

Kahneman Daniel, Riepe Mark W., 1998, Aspects of Investor Psychology, *Journal of Portfolio Management, Vol. 24 Issue 4*, p52-65.

Kahneman Daniel, Tversky Amos, 1979, Prospect Theory: An Analysis of Decision Under Risk, *Econometrica, Vol. 47 Issue 2*, 263-291.

Karni, Edi, 1985, Increasing Risk with State-Dependent Preferences, *Journal of Economic Theory, Vol. 35 Issue 1*, 172-178.

Keating C., Shadwick W.F., A Universal Performance Measure, *Journal of Performance Measurement,.vol.6,.no.3*, 59-84.

Kokkonen Joni, Suominen Matti, 2011, Hedge Funds, Misvaluation, and The HML Factor Returns, Available at SSRN: http://ssrn.com/abstract=1717748 or http://dx.doi.org/10.2139/ssrn.1717748.

Kwag S., Lee S., 2006, Value Investing and the Business Cycle, *Journal of Financial Planning, 19, 1*, 64-71.

Lakonishok J., Lee I., Poteshman A.M., 2004, *Behavioral Finance and the Option Market*, Working Paper, Department of Finance, University of Illinois at Urbana-Champaign.

Lakonishok Josef, Shleifer Andrei, Vishny Robert W., 1994, Contrarian Investment, Extrapolation, and Risk, *Journal of Finance, Vol. 49 Issue 5*, p1541-1578.

Leivo Timo H., Pätäri Eero J., 2011, Enhancement of Value Portfolio Performance Using Momentum and the Long-Short Strategy: The Finnish evidence, *Journal of Asset Management, Vol. 11, Issue 6*, 401-416.

Leung Mark T., Hazen Daouk, Chen An-Sing, 2000, Forecasting Stock Indices: A Comparison of Classification and Level Estimation Models. *International Journal of Forecasting, Vol. 16 Issue 2*, 173-190.

Lev Baruch, Thiagarajan S. Ramu, 1993, Fundamental Information Analysis, *Journal of Accounting Research, Vol. 31 Issue 2*, 190-215.

Liu Jing, Nissim Doron, Thomas Jacob, 2002, Equity Valuation Using Multiples, *Journal of Accounting Research, Vol. 40 Issue 1*, 135-172.

Magnuson Nicholas, 2011, The Role of Expectations in Value and Glamour Stock Returns, *Journal of Behavioral Finance, 2011, Vol. 12 Issue 2*, 98-115.

Mandelbrot Benoit, Hudson Richard L., 2010, *The (Mis) Behaviour of Markets: A Fractal View of Risk, Ruin and Reward*, Profile Books Ltd.

Mehra Rajnish, Prescott Edward C., 1985, The Equity Premium, *Journal of Monetary Economics, Vol. 15 Issue 2*, 145-161.

Mitchell Mark, Pulvino Todd, Stafford Erik, 2002, Limited Arbitrage in Equity Markets, *Journal of Finance, Vol. 57, Issue 2*, 551-584.

Mohanram Partha, 2005, Separating Winners from Losers among Low Book-to-Market Stocks Using Financial Statement Analysis, *Review of Accounting Studies, Vol. 10, Issue 2/3*, 133-170.

Nofsinger John R., Sias Richard W., 1999, Herding and Feedback Trading by Institutional and Individual Investors, *Journal of Finance, Vol. 54, Issue 6*, 2263-2295.

Nofsinger John R., Sias Richard W., 1999, Herding and Feedback Trading by Institutional and Individual Investors, *Journal of Finance, Vol. 54 Issue 6*, 2263-2295.

Odean Terrance, 1998, Volume, Volatility, Price, and Profit When All Traders Are Above Average, *Journal of Finance, Vol. 53, Issue 6*, 1887-1934.

Olson D., Mossman C., 2003, Cross-Correlations and Predictability of Stock Returns, *Journal of Forecasting, Vol. 20 Issue 2*, 145-160.

Paterson Scott, 2010, *The quants: How a new breed of math whizzies conquered the Wall Street and nearly destroyed it*, Crown Business.

Pesaran Hashem, Timmermann Allan, 2005, Real-Time Econometrics, *Econometric Theory, Vol. 21 Issue 1*, 212-231.

Phillips, P.C.B., 2005, Automated Discovery in Econometrics, *Econometric Theory 21(1)*, 3–20.

Piotroski Joseph D., 2000, Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers, *Journal of Accounting Research, Vol. 38 Issue 3*, 1-41.

Polkovnichenko Valery, 2005, Household Portfolio Diversification: A Case for Rank-Dependent Preferences, Review of Financial Studies, Vol. 18 Issue 4, 1467-1502.

Pompian Michale M., 2012, *Behavioral Finance and Wealth Management: How to Build Optimal Portfolios That Account for Investor Biases*, Wiley Finance.

Poterba James M., Summers Lawrence H., James M., 1988, Mean Reversion in Stock Prices: Evidence and Implications, *Journal of Financial Economics, Vol. 22 Issue 1*, p27-59.

Ray C., 2010, *Extreme Risk Management: Revolutionary Approaches to Evaluating and Measuring Risk*, The McGraw-Hill.

Reinganum, Marc R., 1988, The Anatomy of a Stock Market Winner, *Financial Analysts Journal, 44, 2*, 3-16.

Roll Richard, 1977, A Critique of The Asset Pricing Theory's Tests Part I,: On Past and Potential Testability of the Theory, *Journal of Financial Economics, Vol. 4 Issue 2*, 129-176.

Rovine Michael J., von Eye Alexander 1991, *Applied Computational Statistics in Longitudinal Research*, Academic Press

Ryan Thomas P. 2006, *Modern Experimental Design*, Wiley

Scharfstein David S., Stein Jeremy C, Herd behavior and investment, *American Economic Review, Vol. 80 Issue 3*, p465.

Schwager, Jack D. 2012, Hedge Fund Market Wizards: How Winning Traders Win, Wiley.

Schwager, Jack D. 2013, Market Sense and Nonsense: How the Markets Really Work, Wiley.

Scruggs John T., 2007, Noise Trader Risk: Evidence from the Siamese Twins, *Journal of Financial Markets, Vol. 10, Issue 1*, 76-105.

Sheather Simon J. 2009. *A Modern Approach to Regression with R*, Springer
Shefrin Hersh, 2008, *A Behavioral Approach to Asset Pricing 2nd ed.*, Elsevier.

Shefrin Hersh, Statman Meir, 1994, Behavioral Capital Asset Pricing Theory, Journal of Financial & Quantitative Analysis, Vol. 29 Issue 3, 323-349.

Shen Yiyu, Xu Yexiao, 2008, Unique Factors, AFA 2009 San Francisco Meetings Paper. Available at SSRN: http://ssrn.com/abstract=972229 or http://dx.doi.org/10.2139/ssrn.972229.

Sloan Richard G., 2001, Discussion of: "Contextual Fundamental Analysis Through the Prediction of Extreme Returns*", Review of Accounting Studies, Vol. 6 Issue 2/3*, 191-195.

Smith P., Wickens M., 2002, Asset Pricing with Observable Stochastic Discount Factors, *Journal of Economic Surveys, Vol. 16 Issue 3*, 397-446.

Soros George, 1987, *The Alchemy of Finance: Reading the mind of the market*, John Wiley and Sons.

Stoyanov Stoyan V., Racheva-Iotova Borjana, Rachev Svetlozar T., Fabozzi Frank J., 2010, Stochastic Models for Risk Estimation in Volatile Markets: A Survey, *Annals of Operations Research, Vol. 176 Issue 1*, 293-309.

Stoyanov Stoyan V., Racheva-Iotova Borjana, Rachev Svetlozar T., Fabozzi Frank J., 2011, Fat-Tailed Models for Risk Estimation, *Journal of Portfolio Management, Vol. 37 Issue 2*, 107-117.

Swensen D. F. *Pioneering Portfolio Management: An Unconventional Approach to Institutional Investment*, New York: Free Press.

Tabachnick Barbara G, Fidell Linda S 2007. *Using Multivariate Statistics*, Pearson

Triana Pablo, 2009, *Lecturing Birds on Flying – Can Mathematical Theories Destroy the Financial Markets?* John Wiley & Sons.

Trueman Brett; Wong M.H. Franco, Xiao-Jun Zhang, 2000, The Eyeballs Have It: Searching for the Value in Internet Stocks, *Journal of Accounting Research, Vol. 38 Issue 3*, 137-162.

Tufféry Stéphane 2011. *Data Mining and Statistics for Decision Making*, Wiley

Ullah Aman., Giles David E. A., 2011, *Handbook of Empirical Economics and Finance*, CRC Press.

Welch Ivo, 2001,The Equity Premium Consensus Forecast Revisited, Working Papers Yale School of Management's Economics Research Network, p1.

Witten Ian H., Eibe Frank 2005. *Data Mining: Practical Machine Learning Tools and Techniques*, Elsevier

Yamai Yasuhiro, Yoshiba Toshinao, 2002, Value-at-Risk Versus Expected Shortfall: A Practical Perspective, *Journal of Banking & Finance, Vol. 29, Issue 4*, 997-1015.

Yan Zhipeng, Yan Zhao, 2011, When Two Anomalies meet: Post Earnings Announcement Drift and the Value-Glamour Anomaly, *Financial Analysts Journal, Vol. 67, No. 6*, 46-60.

Yu C., 2008, Institutional Trading and Price Momentum, *International Review of Finance. 8(1-2)*, 81-102.

arena Lu, 2005, The Value Premium, *Journal of Finance, Vol. 60 Issue 1*, 67-103.

Zopounidis Constantin, Doumpos Michael, Pardalos Panos M. 2008, *Handbook of Financial Engineering*, Springer.

APPENDIX 1: Model Data

C1N	A2O	A4O	A8O	A13O	C5I	A17O	A21O	A29O	HPR-%	A34O	A35O	C2N	B12I
S0	2	1	0	1	0.876	3	1	0	2	0	2	S1	1
S1	0	0	2	1	0.876	3	2	2	3	2	0	S1	1
S2	0	2	1	1	0.876	3	2	0	2	1	1	S1	1
S3	2	1	0	1	0.876	2	2	2	3	0	2	S1	1
S2	2	2	0	1	0.876	1	2	1	2	0	1	S1	1
S0	2	1	0	1	1.002	3	1	0	1	0	2	S1	2
S0	2	0	0	0	1.002	2	1	1	2	0	0	S4	1
S1	0	0	1	1	1.002	3	2	2	3	2	0	S1	2
S2	0	2	0	0	1.002	3	2	0	2	0	1	S1	2
S2	2	1	0	0	1.002	3	2	1	2	0	0	S1	1
S5	0	2	0	0	1.126	3	2	0	3	0	1	S6	1
S1	0	0	1	1	1.126	3	2	2	3	2	0	S1	3
S2	2	0	0	1	1.126	3	2	0	2	0	0	S4	1
S0	0	1	0	1	1.126	2	1	1	1	0	2	S1	1
S2	0	2	1	0	1.126	3	2	0	2	1	1	S1	3
S2	2	0	0	1	1.016	2	2	0	3	0	2	S4	1
S0	0	0	0	0	1.016	1	2	0	3	0	2	S7	1
S0	2	2	0	0	1.016	2	1	0	2	0	2	S1	1
S0	0	2	2	0	1.016	2	2	0	2	2	1	S7	1
S0	2	2	0	0	1.016	1	1	0	2	0	2	S4	1
S5	0	2	0	0	0.689	3	1	0	3	0	1	S6	2
S2	2	0	0	1	0.689	2	2	0	3	0	2	S4	2
S0	2	2	0	0	0.689	2	1	0	1	0	2	S1	2
S0	0	0	0	0	0.689	2	2	0	3	0	0	S7	1
S0	2	2	0	0	0.689	1	1	0	3	0	2	S4	2
S2	2	0	0	1	0.36	1	0	0	2	0	2	S7	1
S2	2	0	0	1	0.36	2	2	0	3	0	2	S4	2

S0	0	0	0	1	0.36	2	2	0	2	0	0	S7	2
S0	2	2	0	0	0.36	2	1	0	1	0	2	S1	2
S0	0	2	0	0	0.36	2	1	0	2	0	1	S7	2
S2	2	2	0	1	0.05	2	2	0	2	0	2	S4	3
S0	0	0	0	0	0.05	2	2	0	2	0	0	S7	3
S2	2	2	0	0	0.05	2	2	0	2	0	2	S4	1
S0	0	2	2	0	0.05	2	1	0	2	0	1	S7	3
S0	2	2	0	0	0.05	2	1	0	2	0	2	S1	3
S5	0	0	0	1	-0.088	1	2	0	2	2	1	S4	1
S5	0	2	2	2	-0.088	1	1	0	3	2	0	S1	1
S2	2	2	0	0	-0.088	2	2	0	3	0	2	S4	2
S0	0	0	0	1	-0.088	2	2	0	2	0	0	S7	4
S0	0	2	2	0	-0.088	2	2	0	2	0	1	S7	4
S2	2	2	0	0	-0.177	1	2	0	3	0	2	S7	1
S2	2	2	0	0	-0.177	2	2	0	3	0	2	S4	3
S5	2	0	0	1	-0.177	1	2	0	3	0	1	S4	2
S5	0	2	2	2	-0.177	1	1	0	3	2	0	S1	2
S0	0	2	0	0	-0.177	2	1	0	3	0	1	S7	5
S3	0	0	0	0	-0.051	1	2	0	3	0	2	S1	1
S3	0	0	0	0	-0.051	1	2	0	3	2	0	S4	1
S0	0	1	0	1	-0.051	3	2	2	3	0	0	S7	1
S0	0	2	0	0	-0.051	2	1	2	3	0	0	S7	1
S2	1	2	0	0	-0.051	1	1	0	2	0	0	S4	1
S3	0	0	0	0	-0.052	1	2	0	2	0	0	S1	2
S3	1	0	0	0	-0.052	2	2	1	2	0	2	S4	1
S0	0	2	0	0	-0.052	2	1	1	2	0	0	S7	2
S0	0	2	1	0	-0.052	3	2	0	2	0	0	S7	2
S2	1	2	0	0	-0.052	1	1	0	2	0	0	S4	2
S3	0	0	2	0	-0.098	1	2	0	2	2	1	S4	1
S2	0	0	1	0	-0.098	1	2	0	3	0	2	S1	1

S5	2	1	0	1	-0.098	1	2	0	1	0	2	S4	3
S3	0	0	2	0	-0.098	1	1	1	1	2	1	S6	1
S0	0	0	0	0	-0.098	1	1	0	1	0	1	S1	1
S3	0	2	2	0	-0.032	1	1	0	1	2	1	S4	2
S3	0	0	1	0	-0.032	1	1	1	1	2	1	S6	2
S3	0	0	2	0	-0.032	1	2	1	1	2	2	S1	1
S2	0	2	0	0	-0.032	2	2	0	1	0	0	S4	4
S0	2	0	0	0	-0.032	1	1	0	1	0	1	S1	2
S3	0	2	2	0	-0.306	1	1	0	1	2	2	S4	3
S3	0	0	0	0	-0.306	2	2	2	1	0	2	S1	2
S0	2	0	0	0	-0.306	1	2	0	2	0	0	S1	1
S3	0	0	0	0	-0.306	1	1	1	1	2	1	S6	2
S0	2	0	0	0	-0.306	1	1	0	1	0	1	S1	3
S3	0	0	0	0	-0.279	2	2	2	1	0	2	S1	2
S5	2	0	0	0	-0.279	2	1	0	1	0	0	S1	1
S3	1	0	0	1	-0.279	1	1	2	2	2	1	S6	3
S0	2	0	0	0	-0.279	1	1	1	2	0	0	S7	1
S0	2	0	0	0	-0.279	1	1	0	1	0	1	S1	4
S0	2	2	0	2	-0.204	2	0	0	3	0	2	S1	1
S0	0	0	0	0	-0.204	1	1	0	2	0	2	S7	2
S2	0	0	0	1	-0.204	1	1	1	2	0	2	S6	1
S5	0	2	2	0	-0.204	1	1	0	2	0	0	S7	1
S0	2	0	0	0	-0.204	1	1	0	2	0	0	S1	5
S3	2	0	0	0	-0.034	2	1	2	2	0	2	S1	3
S2	2	0	0	1	-0.034	1	1	1	2	0	2	S6	2
S0	2	0	0	0	-0.034	1	1	0	2	0	2	S7	3
S5	0	2	2	0	-0.034	1	1	0	2	1	0	S7	2
S0	0	1	2	0	-0.034	1	1	0	2	0	0	S7	1
S0	2	0	0	0	0.405	1	0	0	2	0	2	S7	4
S2	2	0	0	1	0.405	1	2	1	2	0	2	S6	3

S0	0	1	2	1	0.405	1	1	0	2	2	0	S7	2
S0	2	2	0	0	0.405	2	1	0	2	0	0	S7	6
S0	2	0	0	1	0.405	1	1	1	1	0	0	S1	6
S3	0	2	0	0	0.594	2	1	0	2	0	2	S7	1
S5	2	0	0	1	0.594	2	1	0	2	0	1	S1	1
S3	2	2	0	0	0.594	1	1	0	3	0	0	S1	1
S2	2	0	0	1	0.594	1	2	0	2	0	2	S6	4
S5	0	2	2	0	0.594	1	1	0	2	0	0	S7	3
S2	2	0	0	0	0.791	1	1	0	1	0	0	S1	1
S1	0	0	0	1	0.791	3	2	0	2	2	0	S1	1
S0	0	1	0	2	0.791	3	2	1	2	0	0	S7	3
S0	1	0	0	0	0.791	2	1	1	1	0	1	S1	7
S5	0	2	2	0	0.791	1	1	0	2	1	2	S7	4
S1	0	2	0	1	0.645	3	2	0	2	2	0	S1	2
S0	0	0	0	0	0.645	2	2	2	2	0	2	S6	1
S5	0	1	0	0	0.645	1	1	0	2	0	2	S7	5
S0	0	0	2	1	0.645	2	2	0	2	0	1	S7	1
S2	2	2	0	0	0.645	3	2	0	1	0	2	S4	1
S1	0	2	0	1	0.529	3	2	0	2	2	2	S1	3
S0	2	0	0	0	0.529	2	2	0	1	0	2	S6	2
S0	0	0	2	1	0.529	2	2	0	2	0	1	S7	2
S2	2	2	0	0	0.529	3	2	0	1	0	2	S4	2
S5	0	1	2	0	0.529	1	1	0	2	1	2	S7	6
S1	0	2	0	1	0.537	3	2	0	2	2	2	S1	4
S1	0	0	0	1	0.537	3	1	0	2	0	0	S1	1
S0	0	0	2	1	0.537	2	2	0	1	2	1	S7	3
S0	2	2	0	0	0.537	2	1	0	1	0	2	S6	3
S5	0	0	2	0	0.537	1	1	0	2	1	2	S7	7
S0	0	0	2	0	0.441	3	2	0	2	2	1	S7	4
S0	0	0	2	1	0.441	2	2	0	2	2	2	S7	4

S0	0	0	2	0	0.441	3	2	0	2	2	2	S7	1
S0	2	2	0	1	0.441	2	1	0	2	0	2	S6	4
S0	0	1	2	0	0.441	2	1	0	2	2	1	S7	3
S5	0	0	0	1	0.623	1	2	0	2	0	2	S7	1
S0	2	0	0	0	0.623	1	1	0	2	0	0	S7	5
S0	0	0	2	0	0.623	3	2	1	3	2	2	S7	2
S0	0	1	2	0	0.623	3	1	0	2	0	1	S7	4
S0	0	0	0	0	0.623	3	2	0	2	0	1	S7	5
S3	0	2	0	0	0.647	2	0	0	2	0	2	S7	2
S5	0	0	0	1	0.647	1	2	0	2	0	1	S7	2
S0	0	2	0	1	0.647	1	1	1	2	2	2	S1	1
S0	0	0	2	0	0.647	3	2	1	3	2	2	S7	3
S0	0	0	2	0	0.647	2	1	0	2	0	1	S7	5
S0	0	1	0	1	0.65	1	1	1	2	2	2	S1	2
S0	2	0	0	0	0.65	1	1	0	2	0	0	S7	6
S0	0	1	2	0	0.65	3	1	0	1	1	1	S7	6
S0	0	0	2	1	0.65	3	2	0	3	2	2	S7	4
S3	0	2	2	0	0.65	3	2	2	2	2	2	S1	1
S0	0	0	0	2	0.769	3	2	0	3	0	2	S7	5
S5	0	1	0	2	0.769	1	0	0	1	2	2	S1	1
S0	2	0	0	0	0.769	1	1	0	2	0	1	S7	7
S0	0	1	1	0	0.769	3	1	0	2	0	1	S7	7
S5	0	1	1	0	0.769	1	1	0	2	0	2	S7	8
S3	2	2	1	1	0.765	2	1	2	3	0	2	S7	1
S5	0	0	0	2	0.765	1	0	0	1	2	2	S1	2
S0	0	0	2	0	0.765	3	1	0	3	0	1	S7	8
S2	2	0	0	1	0.765	2	1	2	3	0	1	S7	2
S5	0	0	2	0	0.765	1	1	0	2	1	2	S7	9
S0	0	0	2	1	0.693	3	2	0	3	1	1	S7	1
S0	0	2	0	1	0.693	1	1	2	2	2	2	S1	3

S3	2	2	0	1	0.693	2	1	2	3	0	2	S7	2
S0	0	1	0	0	0.693	3	1	0	3	0	1	S7	9
S5	0	0	2	0	0.693	1	1	0	3	0	2	S7	10
S0	0	0	1	2	0.647	3	2	0	3	0	0	S7	2
S0	0	0	2	1	0.647	3	2	0	2	2	1	S7	6
S0	0	0	2	0	0.647	1	2	2	3	2	2	S1	1
S0	0	1	2	1	0.647	2	1	0	3	0	1	S7	10
S5	0	1	0	0	0.647	1	1	0	3	0	2	S7	11
S0	0	0	2	0	0.569	3	2	0	3	2	1	S7	7
S5	0	0	0	1	0.569	1	2	0	1	1	1	S4	1
S3	2	2	0	0	0.569	2	1	1	3	0	2	S7	3
S0	0	2	0	1	0.569	1	1	1	2	2	0	S6	1
S2	0	2	0	1	0.569	1	1	1	3	0	0	S7	1
S0	0	0	0	0	0.336	3	2	1	3	0	1	S7	8
S0	0	0	2	0	0.336	2	2	0	3	0	1	S7	1
S3	2	2	0	0	0.336	2	1	1	3	0	2	S7	4
S2	0	2	0	1	0.336	1	1	2	3	0	0	S7	2
S0	2	1	0	1	0.336	1	1	0	3	1	2	S4	1
S0	0	0	2	0	0.191	2	2	0	3	2	0	S7	2
S3	2	2	0	0	0.191	1	1	0	3	0	2	S7	5
S0	2	1	0	0	0.191	1	1	1	3	2	2	S1	1
S3	0	0	0	0	0.191	1	1	1	2	0	0	S6	1
S2	0	0	2	1	0.191	1	2	1	3	0	0	S7	1
S3	2	2	0	0	-0.191	1	1	0	3	0	2	S7	6
S0	2	1	0	0	-0.191	1	1	1	2	2	2	S1	2
S3	2	0	0	0	-0.191	2	1	1	1	0	1	S4	2
S3	2	2	0	0	-0.191	1	1	1	2	2	2	S6	1
S0	2	1	0	0	-0.191	1	1	0	3	0	2	S4	2
S0	0	2	0	1	-0.899	1	2	1	2	0	1	S6	1
S0	2	2	0	0	-0.899	1	1	1	2	1	2	S1	3

S3	2	0	0	0	-0.899	2	1	1	1	0	1	S4	3
S0	2	0	0	1	-0.899	1	1	2	1	0	2	S6	1
S2	0	0	0	0	-0.899	1	0	2	2	2	0	S6	1
S1	0	2	0	1	-1.586	3	2	0	1	0	2	S1	1
S2	1	2	0	1	-1.586	1	2	0	2	0	0	S4	1
S3	2	0	0	0	-1.586	2	1	1	1	0	1	S4	4
S0	0	0	0	0	-1.586	1	1	2	1	0	0	S1	1
S0	0	2	0	1	-1.586	1	2	1	1	0	1	S6	2
S0	0	0	2	0	-2.433	1	1	2	1	0	2	S1	1
S3	2	0	0	0	-2.433	3	1	2	1	0	1	S4	5
S2	0	0	0	0	-2.433	1	1	0	2	0	0	S6	1
S0	0	0	2	1	-2.433	1	2	2	1	2	1	S6	3
S3	2	2	0	0	-2.433	1	1	1	1	0	2	S6	2
S3	2	0	0	0	-2.967	1	1	2	3	0	0	S1	1
S0	0	0	2	1	-2.967	1	2	1	1	2	1	S6	4
S2	0	0	1	1	-2.967	1	2	1	1	2	0	S6	1
S0	0	2	2	0	-2.967	2	1	0	3	2	0	S7	11
S3	2	2	0	1	-2.967	1	1	1	1	0	2	S6	3
S0	2	0	0	0	-2.929	2	2	0	2	0	0	S7	1
S3	2	2	0	0	-2.929	1	1	0	3	0	0	S1	2
S0	0	2	2	0	-2.929	2	1	0	3	0	2	S7	12
S0	2	2	0	0	-2.929	1	1	0	1	0	1	S7	1
S3	2	2	0	1	-2.929	1	1	2	3	0	2	S6	4
S3	2	2	0	0	-2.335	1	1	0	2	0	0	S1	3
S2	0	0	1	0	-2.335	1	1	1	2	2	0	S6	2
S0	0	1	2	0	-2.335	2	2	0	3	2	2	S7	13
S0	0	0	0	1	-2.335	2	1	0	2	0	1	S7	3
S0	0	0	1	0	-2.335	1	1	0	1	2	0	S1	1
S0	0	0	1	0	-1.319	1	0	2	1	2	2	S1	1
S2	0	0	0	1	-1.319	1	2	2	2	0	1	S1	1

S0	0	2	2	0	-1.319	2	1	0	3	0	2	S7	14
S0	0	0	2	0	-1.319	2	1	0	1	0	0	S7	4
S0	0	0	0	1	-1.319	1	1	1	1	1	0	S1	2
S0	0	0	0	0	-0.412	1	0	2	1	2	1	S1	2
S0	0	0	1	1	-0.412	2	1	0	1	0	0	S7	5
S2	0	0	0	1	-0.412	1	2	2	2	0	1	S1	2
S2	0	0	1	1	-0.412	1	1	2	3	2	0	S6	3
S0	0	2	0	1	-0.412	2	1	0	2	0	2	S7	15
S0	0	2	0	1	0.337	1	1	2	3	0	0	S4	1
S0	0	0	2	0	0.337	2	1	0	2	0	0	S7	6
S2	0	0	0	1	0.337	1	2	2	3	0	2	S1	3
S2	0	0	1	0	0.337	1	1	1	3	2	0	S6	4
S0	0	2	2	1	0.337	2	1	0	3	0	0	S7	16
S0	0	0	0	0	0.596	2	2	1	3	0	0	S7	7
S0	0	1	2	1	0.596	2	1	2	3	2	0	S7	17
S2	0	2	2	1	0.596	1	2	1	3	2	1	S6	2
S2	0	1	0	1	0.596	1	2	2	3	0	2	S1	4
S0	0	1	0	1	0.596	1	1	2	3	1	0	S1	1
S3	2	2	1	0	0.509	1	1	1	2	1	2	S6	5
S3	0	0	0	1	0.509	3	2	1	2	0	2	S1	1
S2	0	0	1	0	0.509	1	1	0	2	2	0	S6	5
S0	0	0	2	1	0.509	1	1	2	2	2	1	S6	1
S0	0	1	1	1	0.509	1	2	2	2	2	0	S6	5