THÈSE / Télécom Bretagne
sous le sceau de l'Université européenne de Bretagne

pour obtenir le grade de Docteur de Télécom Bretagne
En accréditation conjointe avec l'Ecole doctorale Matisse
Mention : Traitement du Signal et Télécommunications

présentée par
Alya Itani
préparée dans le département Image et traitement de l'information
Laboratoire Labsticc

Reasoning Methods for Merging Financial Technical Indicators

Thèse soutenue le 4 décembre 2014
Devant le jury composé de :

Eloi Bossé
Chercheur, Ecole Polytechnique de Montréal / président

Alain Appriou
Directeur de recherche, Onera - Chatillon / rapporteur

Yves Rakotondratsimba
Professeur, Ecole Centrale d'Electronique – Paris / rapporteur

Ali Hamié
Professeur, Arts, Sciences and Technology University – Liban / Examinateur

Adnan Hamzeh
Professeur, Arts, Sciences and Technology University – Liban / examinateur

Basel Solaiman
Professeur, Télécom Bretagne / examinateur

Jean-Marc Le Caillec
Professeur, Télécom Bretagne / directeur de thèse

No# of Order: 2014telb0348

Sous le sceau de l'Université européenne de Bretagne

TÉLÉCOM BRETAGNE

En accréditation conjointe avec l'Ecole Doctorale Matisse

Co-encadrement avec l'AUL Liban

Ecole Doctorale - MATISSE

Reasoning Methods for Merging Financial Technical Indicators

THÈSE DE DOCTORAT

Mention: Traitement du Signal et Télécommunications

Présentée par ***Alya Itani***
Departement: Image et Traitement de l'Information
Laboratoire: Lab-STICC UMR CNRS 6285

Directeur de thèse: Jean-Marc Le Caillec
Soutenue le 4 Décembre 2014

Jury:

M. Jean-Marc Le Caillec, Professeur à Telecom Bretagne *(Directeur de la thèse)*
M. Basel Solaiman, Professeur à Telecom Bretagne *(Prèsident de jury)*
M. Ali Hamié, Doyen Professeur à l'AUL *(Co-encadrant de la thèse)*
M. Alain Appriou, Directeur de Recherche à l'ONERA
M. Adnan Hamzeh, President Professeur à l'AUL
M. Yves Rakotondratsimba, Professeur à l'Ecole Centrale d'Electronique

"The road to success is not a path you find but a trail you blaze."

Robert Brault

I dedicate this thesis to my parents Ali and Soheir Itani. I hope this achievement brings you honor. Thank you for providing me with the best education, support, and love. I wish nothing from this life other than making you proud.
I would also like to thank my sisters, family and friends for their ongoing encouragement, and my professors for their guidance throughout my reasearch.
I dedicate a special thanks to professors Ali Hamie and Adnan Hamzeh for believing in me and giving me the chance to prove myself.

Alya Itani

Résumé

La gestion de portefeuille en finance devrait consister à faire des bènèfices tout en minimisant le risque. Cependant, la difficulté principale de cette opération réside dans la nature extrêmement volatile des prix des titres sur le marché. Pour cette raison, des techniques d'analyse de titres ont été développées pour aider les gérants de portefeuille à prévoir les changements futurs des prix afin de prendre des décisions pertinentes.

La première technique est l'analyse fondamentale qui est basée sur une étude détaillée des facteurs fondamentaux de l'entreprise émettant le titre. Cette analyse, en plus d'être complexe et consommatrice de temps, est sujette à certaines réticences après le développement de la finance comportementale qui a remis en cause les notions d'efficacité des marchés. Cette remise en cause a suscité l'intérêt des gestionnaires de portefeuille pour un autre type d'analyse, qui existait depuis le dèveloppement des thèories mathèmatiques et des statistiques au dèbut du XX siécle, l'analyse technique. Cette technique repose sur deux hypothèses. La première est que toute l'information disponible sur une entreprise est immédiatement incluse dans le prix de ses titres. La seconde est qu'il existe des faits stylisés qui se retrouvent régulièrement dans les séries chronologiques financières. Depuis ses débuts, cette approche a obtenu de beaux succès, qui en a fait un recours possible des gestionnaires de portefeuilles et des analystes. Elle est aussi le point de départ du travail de thèse.

Afin de créer des bénéfices en gestion de portefeuille, beaucoup de méthodes dérivées de l'intelligence artificielle ont été appliquées au domaine de la finance. En effet, l'analyse technique doit faire face aux incertitudes des séries chronologiques financières, l'influence d'émotions humaines, l'ambiguïté et le manque de précision des données. Afin de prendre en compte dans un raisonnement mathématique les problèmes de gestion de portefeuille, des méthodes issues de l'intelligence artificielle peuvent être utilisèes. Ainsi, des algorithmes de logique floue, des algorithmes génétiques, des réseaux neuronaux, des possibilités et des approches neuronales confuses ont été développés pour la gestion de portefeuille.

L'examen des forces et des faiblesses de chaque approche suggère l'application d'une approche hybride profitant des théories des probabilités et des possibilités. En effet, les probabilités permettent l'apprentissage de données statistiques, autrement dit de prendre en compte l'existence de faits stylisés comme nous l'avons mentionné auparavant et la théorie de possibilité permet de modéliser l'incertitude afin d'incorporer les facteurs humains dans le traitement des données impliquèes dans le processus de décision.

Notre démarche a été de montrer que la fusion de plusieurs indicateurs techniques peut conduire à de meilleures décisions que celles basées sur un indicateur seul afin de prédire les variations de prix et de tendance et donc de prendre une décision dáchat ou de vente pertinente. Nous avons proposé plusieurs systèmes de décisions hybrides pour effectuer la fusion. Grâce à l'estimation de l'entropie via la divergence de Kullback Leibler ainsi que des techniques de transformation de Dubois-Prade, nous avons incorporé dans notre système de fusion un coefficient prenant en compte la fiabilité des indicateurs dans le processus de fusion.

Les systèmes définis ont été testés de manière exhaustive, transparente, le plus précisément possible et ont montré des résultats prometteurs. En particulier, la possibilité d'obtenir de meilleurs résultats, c'est-à-dire de meilleures décisions par fusion, que pour un indicateur a été montrée. De plus, l'inclusion de la fiabilité des décisions sous une forme évolutive au cours du temps est une contribution nouvelle ajoutée aux connaissances disponibles. Une approche par fusion bayésienne a aussi été testée, afin de comparer les résultats de ces réseaux avec ceux obtenus par approche possibiliste.

Ce travail de thèse ouvre la voie à de futurs travaux sur des systèmes de décision dans le domaine financier. Un exemple possible d'extension serait lápplication du système de décision, qui ná été testé que sur des indices, à la gestion complète d'un portefeuille, c'est-à-dire pour toutes les actions négociées en incluant une pondération de type Markowitz. Pour conclure, nous croyons que l'application des méthodes de raisonnement approximatif pour le domaine de la finance est une mine de recherche qui ouvre des horizons illimités.

Abstract

Portfolio management is a mean of making profit and expanding wealth through following different security trading strategies, such as the act of buying a financial security at a certain price, and selling it later at a higher price to make profit out of this trade. However, the main difficulty lies in the diligently varying nature of security prices in the financial market. For that reason, market and security analysis techniques were developed, in order to help traders forecast the future price change in order to make winning decisions.

The first techniques known as fundamental analysis is based on a detailed study of the fundamental factors of the issuing company of the security of interest. Besides being very complex and time consuming this analysis technique became liable to doubts after the arrival of behavioral finance that caused a controversy around its underlying assumptions of market efficiency. This event sequence contributes in directing the interest of traders towards another security analysis technique known as technical analysis. This technique relies on two assumptions, the first is the belief that price immediately integrates all the available information of a security, the second is following the concept of history repeating itself. Ever since its existence, this technique has shown great success, which made it the resort of all financial traders and analysts and the interest of this thesis work motivation.

For the best intention of making money, many reasoning methods have been integrated with finance to help best meet that goal. The studied situation of financial market and technical analysis comes with a big deal of uncertainty, incorporation of human emotions, ambiguity and vagueness. Normally for a mathematical reasoning with the above mentioned challenges it would be convenient to use theories used are reasoning methods of artificial intelligence techniques. Digging into the history of applied methods on studying financial markets, we explore the use of fuzzy logic, genetic algorithms, neural networks, probability, and fuzzy neural approaches. Examining the strengths and weaknesses of each available approach led the orientation of this research into applying a hybrid approach that takes advantage of both theories of probability

and possibility. Since, probability theory is known for its power with learning statistical data, and possibility theory is known for its competences in handling uncertainty and processing any human factor incorporation.

The challenge dwells in proving the superiority of fusing multiple technical indicators over using individual indicators to foresee price change and make a winning decision. This approach includes multiple probability-possibility decision support systems used for studying the effects of multiple indicators fusion on the risk and revenue upon making a trading decision. By that, taking advantage of the Kullback Leibler divergence and Dubois-Prade transformation techniques to provide each indicator with a weight factor that represents its efficiency.

The applied systems have been tested exhaustively and transparently in the most accurate manner, and have shown promising results, thus complying with the main goal of this thesis. The main challenge of superiority of merged indicators over individual ones was achieved in many of the proposed decision support systems. Furthermore, the inclusion of reliability under its dynamic form is proposed as an innovative contribution added to the current available body of knowledge. Also a basic Bayesian fusion approach was introduced that uses learning with Bayesian networks as a means for decision making. The purpose of the BN approach was to compare its outcome to the contributed systems in order to diversify the testing as much as possible.

This proposed work plan is just the door to many potential financial decision support systems. An example of possible extensions, would be an integration of the system with a Marcowitz portfolio allocation system that makes use of the decision output of the hybrid probability possibility decision fusion approach. Finally, we believe that the field of reasoning methods application in finance could be considered as a mine of research, where lies an unlimited horizon of innovation.

Contents

Introduction

0.1 Overview of the Situation Understudy

Trading is defined as *"A basic economic concept that involves multiple parties participating in the voluntary negotiation and then the exchange of one's goods and services for desired goods and services that someone else possesses"*. It is just like any kind of investment acting as a mean for making profit and expanding wealth. However, there are two types of investment that people seek for revenue. The first includes, investment in cars, houses, land and many other real assets, known by real investment. And, the other includes investing in stocks, bonds, treasury bills and other financial securities, it is know by financial investment which is what this work motivation is interested in.

An important fact about financial investment, is that it depends much on price change. Just like any trade, it follows the goal of buying low and selling high to make revenue. Nevertheless, in the world of financial markets, security prices change on daily basis and even less. Security prices in the market change diligently, making it hard for financial traders to take granted winning decisions. For that purpose, we find investors and traders very eager to forecast price movements before occurring in order to guaranty a winning decision at the right time. Many analysis techniques were developed to help traders achieve their goal. The two main techniques adopted are fundamental and technical analysis.

Fundamental analysis relies on a detailed complex examination of the fundamental factors of the underlying company to a security, in order to estimate the actual worth of the issued securities of this company before it is reflected in the market, uncovering by that the companies with valuable assets. This technique, as any other, has some limitations related to time consumption, subjectivity and accuracy. However the main weakness of fundamental analysis lies in its underlying assumption of market efficiency. Behavioral finance has proven the irrationality of markets, due to the incorporation of human psychology and emotions with trading and its direct effect on financial markets, in particular at times of crisis.

This market irrationality theory contributed in directing the interest of traders towards another security analysis technique known as technical analysis. This analysis technique states that prices discount all security fundamentals discarding by that the need of fundamental analysis. Another main assumption of technical analysis is that history repeats itself, thus one can predict the future through analyzing its past behavior. The analysis tools of this technique are normally studies applied on price trends, charts and patterns, with the help of technical indicators. Technical indicators are a set of mathematical formulas applied to the price and volume, to simplify the analysis for traders. In this specific corner of investment lies our thesis motivation for research.

The main challenge of this research is studying whether the integration of multiple indicators and merging their effect in a decision support system would be more profitable than following decisions derived from one indicator. This necessitates a deep understanding of technical analysis, in order to find which reasoning method is best used for achieving this goal. Technical analysis comes with a great deal of uncertainty, ambiguity and vagueness, due to the dependence of its success on many interfering factors. Such factors are, the change of indicator efficiency when applied on a different stocks or on different time horizon of the same stock, the selection of indicators, the parameters used for each indicator, and the human factor integrated with the indicator interpretation process. The success of an indicator forecast can never be granted, even the elite of experts cannot assure the success of a certain followed analysis technique. Yet, if efficiently handled, this analysis can give surprising revenue, which is the reason behind calling it the voodoo of market trading.

The second facing challenge would be choosing the most suitable reasoning method that is capable of efficiently mimicking the analysis of human expert. This is achieved through merging the effect of the most robust indicators.

0.2 Motivation and Contribution

The first section addressed the challenging problems with financial markets and technical analysis, now we highlight the solutions proposed and the contribution of this thesis research on dealing with such challenges. Throughout history many artificial intelligence reasoning methods have been integrated with finance in order to help traders achieve the goal of making maximum profit with minimum risk of loss. Such methods included neural networks, fuzzy logic, probability, genetic algorithms, machine learning and pattern recognition.

A known challenging subject widely handled in the domain of technical analysis is detecting visual technical patterns that closely mimics the recognition of a human expert. And, many innovative approaches have been applied on the matter with relieving inevitable achievements. However, the uncertainty in the financial market is beyond patterns. It is correlated with security selection, indicators efficiency and parameter selection, market efficiency and rationality of traders, the decision making process itself and many other aspects not to be neglected as well.

This motivates the current research into digging more on the available body of knowledge. Other addressed challenges and used reasoning methods include, taking advantage of the interpretation and uncertainty handling competences of fuzzy logic and genetic algorithms for stock selection and deployment of technical trading rules, along its benefits in modeling experts knowledge and its attempts with price evaluation. This technique has overcome its rivals of artificial intelligence reasoning methods, such as neural network which has many proven limitations when handling decision making using technical analysis. One of which is, its inability of handling uncertainty and ambiguity fairly contributing in such environment. However, there exists in the financial markets the ability of using historical data prices of many well known and traded securities. Taking advantage of this added statistical knowledge in the learning and evaluating stocks could give more promising results than exclusively using fuzzy logic on its own. Therefore, we adopt this point of view in the following work motivation. Another potentially beneficial detail is the search of a different analysis technique than fuzzy inference systems and genetic algorithm, that could give importance to one side of information knowledge, and discarding another (a complete expression of trading rules could be very complex and probably impossible to reach).

The fact that all reasoning methods have their limitations drove the motivation of researchers towards the world of hybrid intelligent systems, where many integrated learning and adaptation techniques have been proposed to achieve synergistic effects, such as neuro-fuzzy frameworks and probabilistic-fuzzy systems. Also, the applied methods recorded great contributions to the already available models. Using the statistical powers of probability theory to deal with historical data available for securities of the financial market is definitely a winning added step. However most of the systems are fuzzy rule based systems. The accurate and sufficient modeling of such systems tend to be complex. Furthermore, most of the applied studies do not include the effect of various indicators, excluding by that the benefit of including the widest financial knowledge possible.

Therefore, this work motivation also follows the concept of a hybrid artificial intelligence system to take advantage of the most possible advantageous powers of available paradigms that deal with uncertainty in its various available types in this particular environment under study. In this manuscript, we propose a system that takes advantages of probability theory in dealing with statistical historical data, possibility theory competences in handling uncertainty and dealing with the available human factor, and the foreseeing capabilities of technical analysis with merging information from various technical indicators in the most efficient manner possible. Another applied ans tested system is a basic pure probability fusion system using Bayesian Networks for decision making, in order to compare the resulting outcome of both proposed approaches.

0.3 Outline of the Dissertation

This thesis dissertation is divided into five chapters. Following this introduction is the first chapter of portfolio management. This chapter introduces investment in its available forms, along with offering a detailed explanation of financial markets and different categories of traded security types. The chapter then addresses briefly the two main types of security analysis and continues into presenting portfolio management with its different analysis and selection means, pricing models, and evaluation techniques. The importance of this chapter resides in familiarizing the reader with the basic notions of finance, in order to deliver a better understanding of the situation under which the contribution takes place.

The second chapter of the dissertation delivers a detailed definition the two main techniques of security analysis, besides giving a deep look into fundamental analysis different factors and evaluation techniques. Another interesting part of the chapter is the examination of technical analysis its different price fields, chart types and patterns, accompanied by a listing of the most reputed technical indicators that passed the test of time. In this chapter the importance lies in well understanding technical indicators, where they represent the core of this thesis work motivation.

After covering the needed knowledge on security analysis techniques, the controversy around fundamental knowledge, and the explained drift of trader interests towards technical analysis techniques, chapter three introduces a state of the art and explains the path and orientation of the situation challenges and problems. This chapter gives an accurate introduction to the usual theories and assumptions logically explained to help the reader intuitively and clearly understand the reasons behind the proposed solutions and approaches detailed in the following chapters. Then, the second part of this chapter puts forward a detailed description on the history of reasoning methods and their integration with finance along different aspects. This chapter forms simply the reflection of the challenges and problems of the environment under study and its tools and conditions.

Coming to the fourth chapter, it is necessary to state before proceeding that this is considered the most important chapter of this dissertation, since it includes the contribution of this research motivation. It first introduces the pre-processing general approach which is a hybrid probability-possibility preparatory system, that simply processes data into its fusion-ready state. The chapter then introduces multiple fusion approaches that are fed up with the output data of the pre-processing system, aiming to overcome in performance individual indicators based analysis, and other applied fusion techniques. Followed by a detailed description to the complete testing mechanism used for the sake of consistency of the scientific research.

Finally, chapter four aims to strengthen previous proposed approaches through diversifying the testing strategies, to be as complete and convincing as possible. For that purpose, the chapter introduces a different purely probabilistic fusion approach with Bayesian Networks. The first part of the chapter tackles the basics of Bayesian networks and its different inference techniques, with also stating the different developed algorithms for structure and parameter learning. The second part includes a basic application of the theory for technical indicators fusion with Bayesian Networks with the bnlearn package of the R statistical environment.

Finally, the dissertations is finalized with a section that summarizes all derived conclusions and consequences with a declaration of a fair personal perspective on the applied studies.

CHAPTER 1 Portfolio Management

1.1 Introduction

Generally, an investment comes in different activities and regardless of the means, investment activities try to share a common goal. Investments basically aim to make profit and expand investors wealth, through employing money for a certain period of time, under an acceptable risk for the investor. It could be regarded as simply postponing the desire of revenue for a future period of time in the purpose of its increase, where individuals, corporate sectors, or business organizations engage their savings in income generating assets. However, there are two different types of investment, real and financial investments. The real investment usually deals with actual assets as: cars, houses, lands, etc. while financial investment involves contract forms, such as treasury bills, stocks, bonds, etc. The interest orientation of this thesis concerns the financial investment process where lies the key concepts of financial securities, their analysis techniques and decision making process in the substantially wider context [56]. All financial investments include two very important factors the risk and the return. The return is always accompanied with risk in a way that they depend on each other whenever one chooses to invest with less risk, it is typical for the return decrease along. In order to avoid risk of loosing everything at once, experts developed the idea of diversification where instead of investing in one asset or multiple assets of one type, one can invest in a portfolio of assets with a risk-return combination selected to meet each investor's capabilities and objectives in particular reducing the risk. The second part of this thesis introduces portfolio analysis, and selection techniques, discussing the most used techniques for market portfolio selection and finally discusses some portfolio evaluation techniques.

1.2 Financial Investment

It is widely common that the economic performance, whether growth or crisis, is highly influenced by its financial systems. It is put into belief as times passes that the development of any nation will indeed depend on its financial architecture. This architecture functions through fund allocation of certain entities to other potentially more experienced and productive entities as a way of funds investment. Nevertheless, investment of any form includes risk factors, while the dynamics of the economic system include multiple opportunities and techniques for some good investments to take place. Capital appreciation and noted income could be assured with a well planned investment strategy. Usually, when investors trade in the financial market, they expect a regular flow of income paid out of this process. However, many factors interfere to this formula, and should be well considered and processed.

- **Return** is simply the prize resulting from an investment. The main aim of an investment is collecting return. It could be direct in the form of regular income, or in the form of capital appreciation, where the price or value of the owned asset increases. Return is usually always accompanied by risk, where the higher return gets the higher risk chases it[7].

- **Risk** is normally the danger or harm following the return. it can be described as the disappointment of meeting an expected profit. It can be measured through estimating the difference between the actual return and the expected return. Generally, multiple factors might induce risk, some of which are controllable and some not. However, every investor aims to decrease risk and to increase return in all feasible ways.

- **Safety** is one of the most important factors in a financial investment, mainly investment is employing money in a selected type of financial asset, which differs in regulation from one another. Investment places such as bank deposits, government bonds, equity shares, and so on are considered with little risk, while the obtained return is relatively low.

- **Liquidity** Liquid financial assets provide facility in trading, where it enables investors to cash their investment whenever they find it necessary. This characteristic of flexibility drew the attention of various investors through out time.

- **Hedge** It is commonly used to cut down possible losses and gains that might be obtained by a companion investment. It can be built from various instrument types such as, Swaps, Stocks, Options, Futures etc...

1.3 Financial Securities

Financial securities could be thought of as legal contracts with pre-set conditions, providing investors with rights to receive future benefits. There are multiple ways where securities can be categorized, this chapter limits the categorization of securities into those traded in organized markets [32].

1.3.1 Direct Investment Securities

The securities in this category could be classified according to the investment time horizon used. A illustrated in Figure 1.1 direct investment securities can be divided into these categories:

- **Money Market Securities**, are the type of securities known for having one year or less maturity periods when issued. They are considered short-term securities that are sold by financial institutions, corporations, and governments:
  - ***Treasury Bills***, are known for being the least risky securities to invest in. They are close to being almost risk-free investments. They have short-term maturities, the return they generate is known, and they are usually traded in active markets.

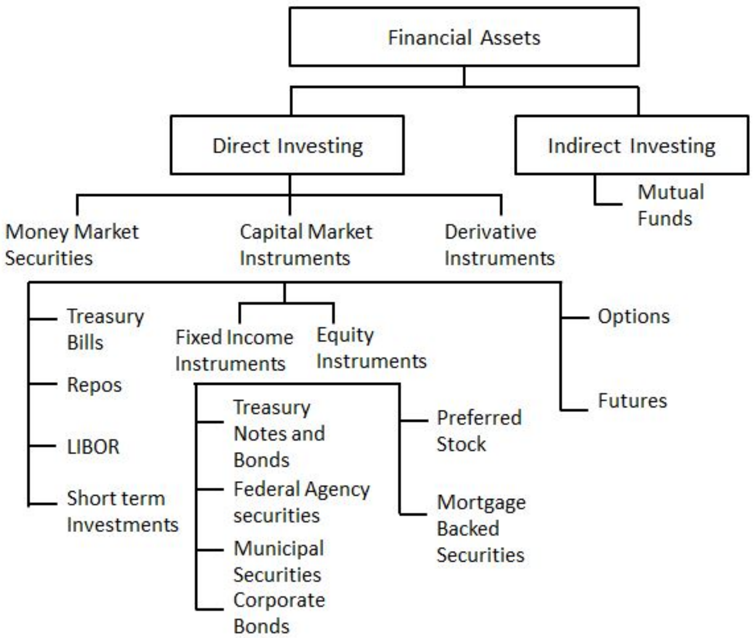


Figure 1.1: Schematic Diagram of Financial Securities Types

  - ***Repurchase Agreements (Repos)***, are agreements of sell and repurchase, with very short maturities not exceeding 14 days. The agreement is usually arranged between a borrower and a lender to trade government securities. The investment takes place as a borrower signs a contract of selling the security to a lender, and buy it back after a certain period of time at a specified price. The return to the lender would be the difference of the two prices.
  - ***The London Interbank Offered Rate (LIBOR)***, is the rate at which London international banks lend each other money as loans. It has the characteristics of long term securities, since the periodic change in its rate, which is not common with short-term traded securities.
  - ***Other Short Term Securities***. Even though securities with short-term maturity are less risky than any other securities, the return generated by their investment is considered evidently low compared to more risky securities. There exist many other short-term securities, such as the negotiable certificates of deposit or also Certificate of deposite (CD), which are certain bank deposits with specified periods of time. Another short term security is the Bankers' acceptances, which are contracts made by banks to pay a certain sum of money on a specific date. The selling rates of both securities depend on the banks that back them.

- **Capital Market Securities**, these types of securities are characterized with having maturity periods more than one year, and some times with no specified maturity. The subcategories belonging to this type of securities are distributed according to the way of paid profit, whether it is a running cash flow over time, or a participation in the future profit of a company.

  - ***Fixed Income Securities***, the specification in these securities is their stated payment schedules, whether usual traditional bonds, or a promised stated amount and date of payment. Following are some examples of fixed income securities: following:

    * *Treasury Notes and Bonds* are simple debt instruments issued by the government, and have maturity periods between one and ten years. Both treasury notes and bonds pay the investor an interest twice a year, and principle price at the end of maturity period.
    * *Federal Agency Securities* are issued by certain federal agencies that have been granted the authority to emit dept when needed and help certain sectors in the economy.

    * *Municipal Securities* are bonds that are also dept instruments, they are issued by entities other than the government, such as cities, school districts, states.
    * *Corporate Bonds* are similar in payment method to government bonds, but these bonds are issued by business entities, that give a higher risk factor than government issued bonds.
  - ***Equity Instruments*** these instruments are known for having a relatively high variability is cash flow received by the investor.
    * *Preferred Stock* are close in concept to life bonds and pay the holder on regular time periods, the difference is that payments are paid as dividends instead of interest.
    * *Mortgage-Backed Securities* are low in risk securities since they are backed by the government, for this reason possibly they are categorized under fixed income securities. They represent a share in a group of mortgages.

- **Derivative Instruments** could be though of as securities whose values are derived from that of the underlying group of securities, for that purpose they are also called contingent securities.
  - ***Options*** are certain kind of securities that gives the holder the right to buy or sell a certain multiple or single securities, while stating the date or period of time of taking action, and specifying a price in advance.
  - ***Futures*** are delayed actions on a security, they are obligations to buy or sell a single or multiple securities with predetermined price and time.

1.3.2 Indirect Investment Securities

The case where an investor directly invests through buying and selling instruments has been discussed above briefly. Another type of investment exists, which is the indirect Investment. Indirect investment is when traders invest in an indirect manner through purchasing shares of investment companies.

- **Mutual Funds**, the concept of the mutual fund is based on holding a securities portfolio with a specified objective and policy. It gives to the investors an opportunity to access diversified portfolios of stocks, equities, bonds, money markets, and other securities.

Now that the most traded types of securities are introduced, it is important to gain a certain familiarity with types of financial markets.

1.4 Financial Markets

Financial markets are the centers of economic development or regression of a nation. It is the place where the different previously described securities are traded, and where borrowers and suppliers of funds meet. Suppliers are normally the parties that supply funds as an investment, whether it is individual investors, companies, firms, or corporations. Any party with a surplus fund dedicated for investment is a supplier. While borrowers simply (as their name indicated), are parties that borrow money from suppliers in different terms, as dept or investment loans, to professionally employ these funds for the intention of making profit. The main and more important factor of markets is the reputation. It directly and greatly affects the allocations of funds. As in financial securities there are different possibilities of categorizing markets. However the general categorization of financial markets is into two types, primary and secondary markets.

1.4.1 Primary Market

Also known as the Initial Public Offer, or the New Issue Market. This market involves new issues of securities, where they are initially sold. Hence, this market provides a direct flow of cash to the party issuing securities. It holds the burden of selling securities to the public. It acts as a main reservoir of funds raised from many entities like individual investors, financial companies, institutions, etc. This makes it the best place for corporate sectors to raise their funds. The role of a Primary market can be resumed into, investigation, underwriting and Distribution.

1.4.2 Secondary Market

The secondary market is specially meant for long-term securities. It is the place that resales the primary market already issued securities. This makes the effect between the two markets inevitable, a strong action on the secondary market causes high demand of new shares issued through the primary market. Thus, growth of the primary market is greatly related to that of the secondary market. Stock Exchanges are the most known secondary markets, whose purpose is controlling, regulating, and assisting the business in trading the securities.

1.5 Security Analysis

Before introducing security analysis and its techniques, one should take a wider look on how it is related to portfolio management, referring to figure 1.2

An Investor mainly aims to take advantage of the fluctuation of security prices, buying at low prices and selling at higher prices to make profit. Nevertheless, profit is always chased by risk in financial investments. The best investors are those able to avoid risk through analyzing securities efficiently, thus choosing the best times to enter and exit the market at the right times with the right shares. To achieve that traders resorted to examining the actual worth of shares and figure out the intrinsic values of securities, this is known as security analysis. With security analysis traders evaluate the price of a security to judge whether it is over or under priced, hence estimate whether entering the trade would lead to gain or loss. There are three main analysis techniques developed throughout time, efficient market theory, fundamental analysis and technical analysis.

1.5.1 Fundamental Analysis

Fundamental analysis refers to studying the real value of a security through examining the efficiency of its underlying company. This necessitates a deep investigation of all fundamental factors that might affect the performance of a company. The studied factors normally include profit margins, balance sheets, growth potential, management strategy, etc. Therefore, mainly any issue that might have a direct or indirect effect

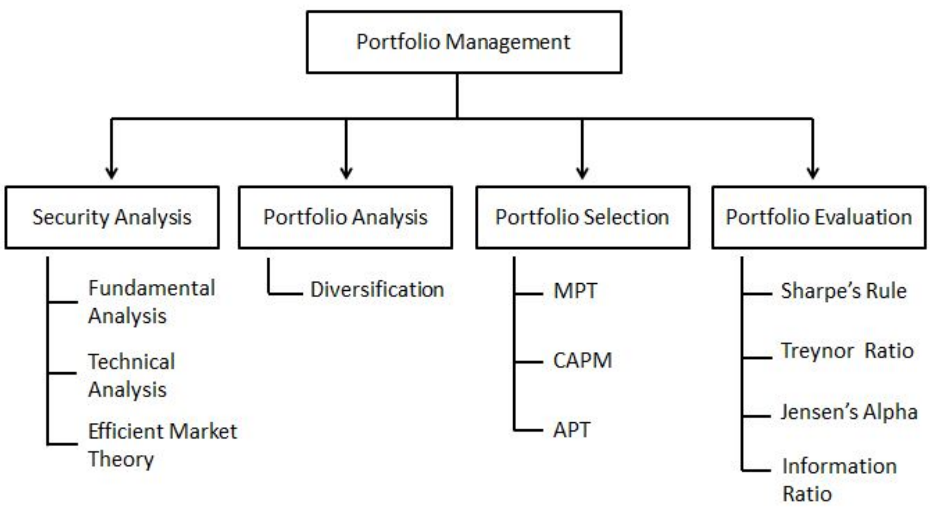


Figure 1.2: Schematic Diagram of the Protfolio Management Process

on a company, is of great importance for fundamental analysts to forecast the future price movement of securities.

1.5.2 Technical Analysis

Technical analysis is another way of security analysis that cares less about qualitative fundamentals, and more about prices, charts and patterns to forecast future price movement. It resorts to analyzing historical price data believing that the future could be derived from studying the past. It uses technical indicators, that are mathematical formulas applied to the price, to study patterns, trends, and other price factors and accordingly make investment decisions. The accumulating evidence of market inefficiency caused a revival of academic interest in technical analysis claims. Since then, it has been showing great predictive power compared to other strategies and analysis.

1.5.3 Efficient Market Theory

This theory, also known as the Efficient Market Hypothesis, asserts that prices are only reflections to current situations, being affected by recent information. This hypothesis mainly comes from the assumption that markets are dynamic and are greatly influenced by multiple factors in the political, economic and business environments. It asserts that markets are informationally efficient since information is considered as a powerful tool for experts. Information flows directly to the market reflecting price changes, and putting markets into vast competitive states. This evident effect of information on the market leaves traders with consciousness, awaiting any change that might affect their traded securities to take immediate actions.
There are three known types of the efficient market hypotheses, weak, semi-strong, and strong. The weak hypothesis claims that prices of traded assets already reflect all past interfering information. The semi-strong hypothesis affirms that prices reflect all available information to the public and that in case of new information availability price will re-change to reflect the new condition. As for the strong hypothesis it asserts that prices directly reflect current information even the not much spread or hidden information. Although the hypotheses attracted experts interest when first developed, a vital ongoing development on behavioral finance which interprets finance from a wider perspective examining the psychology and sociology side of traders, has lead to skepticism and contradiction to the assumptions of efficient market theory. Chapter 2 details more the concept of this theory and its depreciation with the arrival of behavioral finance.

1.6 Portfolio Analysis

It is evident that the financial markets are highly volatile and much aggregated with risk, that encouraged the development on the notion of portfolio. The portfolio is a combination of assets subject to investment. It targets the goal of minimizing risk and maximizing return through diversification while investing. Investors usually have the full choice of which assets to engage and at what manner, they can manipulate their invested assets the way that suits them. Therefore, any investor would normally consider trying to invest in different sets of securities to increase profit.

1.6.1 Diversification

Investing in more than one security is called diversification, it simply refers to diversifying ones investment in different securities to avoid risk of losing everything and payoff from distributed return of multiple securities instead of a single one. Very close to the common say of not putting ones eggs all in the same basket, to avoid complete loss in any case of failure. An important rule in diversification is choosing a portfolio of assets that do not move in perfect unison. In other words it is about choosing not so related assets or assets from different sectors. Taking as example investing in a portfolio of real estate and food and beverage, if some animal disease phenomena occurs, then one could lose in the food and beverage invested securities, but the real estate related securities stay unaffected, this symbolizes the mechanics of diversification.

1.7 Portfolio Selection

A good portfolio is not just a random chosen pool of invested securities, the effectiveness of the portfolio relies on many factors. It is a balanced whole that provides the investor with opportunities and protection through all probabilities. The portfolio selection depends mainly on criteria related to investors aims and needs. Through out time many techniques where developed to serve investors in that purpose. We will discuss in this chapter the most used techniques among financial investors.

1.7.1 Modern Portfolio Theory (MPT)

MPT is an investment theory which aims to maximize portfolio expected return for a given amount of portfolio risk, or equivalently to minimize risk for a given level of

expected return, by carefully choosing the proportions of multiple assets in a portfolio. It uses the concept of diversification and choosing portfolio assets that have collectively less risk when traded together than individually.
Normally, assets of different types move differently. Taking example stock market and bond market securities, they usually move in different directions, ie. they are negatively correlated, thus reducing the risk of portfolios made of both. However, MPT also asserts that risk can be reduced even in the case where assets are not negatively correlated (in fact even if they are positively correlated).
MPT was first introduced in the 1952 by Harry Markowitz. It symbolizes a great evolution in the mathematical modeling of finance. Nevertheless, as behavioral finance started showing great evidence disproving the rationality of markets, many criticisms were addressed to the theory for its basic assumption of markets rationality. Another skeptical aspect in MPT was that it assumes that returns follow a Gaussian distribution, which is in a contradiction with financial experts assuring that returns do not, in any way, follow any symmetric distribution.

MPT lies on the assumption that investors are risk averse, asserting that an investor always chooses the less risky portfolio option. Nevertheless, whenever investors seek higher returns, they must expect a higher risk as well, depending, by that, on the objectives and surplus funds of different investors. MPT models returns as normally or elliptically distributed functions and models risk as the standard deviation of return. Whereas returns in portfolios become the weighted return of constituent securities returns. The mathematical representation is as follows:

- *Portfolio Expected Return*

$$E(R_p) = \sum_i w_i E(R_i) \tag{1.1}$$

  Where $E(R_p)$ is the expected return on the portfolio, $E(R_i)$ is the expected return on asset i, and w_i is the weight of individual asset i estimated by the likely profits of each asset class.

- *Portfolio Return Variance*

$$\sigma_p^2 = \sum_i \sum_j w_i w_j \sigma_i \sigma_j \rho_{ij} \tag{1.2}$$

  Where σ_p^2 is the expected value of the squared deviation of the return on the portfolio from the mean return on the portfolio, w_i and $w_j$are the weights of assets i and j respectively and σ_j and σ_j are the standard deviation of assets i and j respectively representing assets risk. ρ_{ij} is the correlation coefficient between returns of assets i and j defined as $\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$, with $\rho_{ij} = 1$ when $i = j$. Note that dividing by the product of the standard deviation does not change the co-variance. It simply scales it to have values between -1 and +1.

For clarification, we take as example a portfolio of two securities A and B the mathematical representation of portfolio's return and risk would be as follows:

- *Portfolio Return*

$$E(R_p) = w_A E(R_A) + w_B E(R_B) = w_A E(R_A) + (1 - w_A) E(R_B) \tag{1.3}$$

- *Portfolio variance*

$$\sigma_p^2 = w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2 w_A w_B \sigma_A \sigma_B \rho_{AB} \tag{1.4}$$

In financial terms these formulas represent the return and the risk (represented by standard deviation, which is the square root of the variance $\sigma_p = \sqrt{\sigma_p^2}$). Estimating the return and variance can help the investor to choose a portfolio that exactly meets his needs or fittings of return and risk proportions. Where, the interesting case of the portfolio variance formula is when the correlation coefficient is negative ($\rho_{AB} < 0$). Having a negative correlation coefficient according to the formula decreases the portfolio variance which signifies a reduced portfolio risk. Calculating the return and variance of a portfolio is also very important for the diversification strategy as explained in the following section.

MPT Diversification & Efficient Frontier

Diversification in modern portfolio theory is applied by choosing instruments that are not perfectly positively correlated, meaning that the correlation coefficient $-1 < \rho_{ij} < 1$

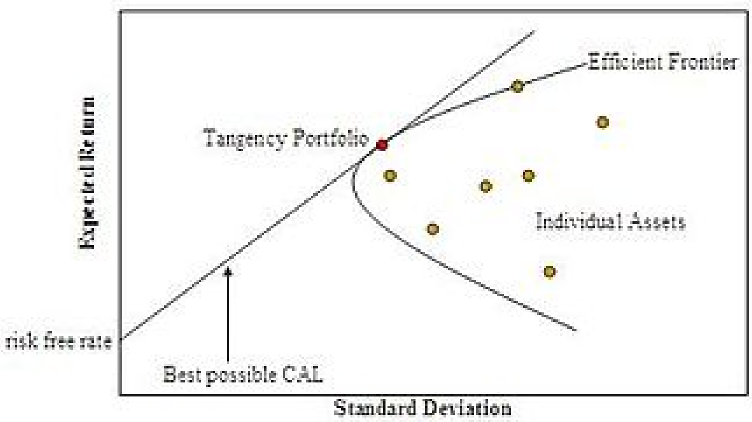


Figure 1.3: The Efficient Frontier Hyperbola of Typical Risky Portfolios

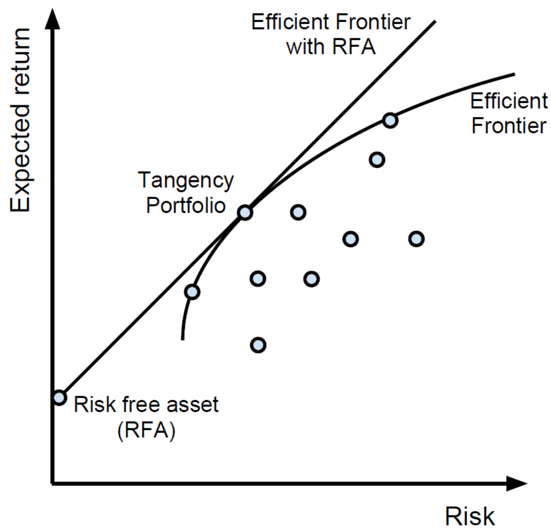


Figure 1.4: The Efficient Frontier Hyperbola of Portfolio with risk-free Asset

should not be positive to be less risky. Thus, by that when diversifying investment one can avoid exposure to individual instruments risk.
We have discussed the trade-off between risk and return always present in all investments, saying that MPT claims maximizing return for a certain level of risk, or vice versa according to investors needs. Thus, an investor must choose the return and risk values that resemble to his preference, to analyze the effect of changing parameters. For that purpose, Markowitz coined the Efficient Frontier which is an MPT concept that shows to investors the best possible return to expect from their portfolio given a certain level of risk, or which is known as the optimal portfolio. It can be described as an upper-half hyperbola between risk and return showing all possible portfolios of all possible levels of risk and return, with the most efficient portfolios lying on the envelope of the hyperbola (figure 1.3). Efficient portfolios are portfolios with the highest risk-return ratios given any parameter value. Meaning that it is always possible to select a portfolio in the efficient frontier that dominates any non-efficient portfolio. The mathematical representation of an efficient frontier is obtained through minimizing the following expression 1.5 given s risk tolerance variable q

$$w^T \Sigma w - qR^T w \tag{1.5}$$

Where w is the vector of portfolio weights, Σ is the co-variance matrix of individual assets forming the portfolio, $q \in [0, \infty)$ is a risk tolerance factor, and $R^T w$ is the portfolio expected return [52].

Portfolios with Risk Free Asset

The risk free asset, as its name indicates, is free of risk which means that it has zero variance and therefore uncorrelated to any other asset. Hence, using a risk free asset along with other assets would result in linearly related changes or return and risk, with just varying proportions. Figure 1.4 illustrates the efficient frontier. We can notice the point of tangency between the hyperbola and the half-line, this points horizontal intercept represents a portfolio fully formed of risk free assets, while the tangency point represents a portfolio without any risk-free asset. The efficient half-line is known as the capital allocation line CAL, estimated by the following equation [32].

$$E(R_C) = R_F + \sigma_C \frac{E(R_P) - R_F}{\sigma_P} \tag{1.6}$$

Where P represents the portfolio of risky assets at the point of tangency with the efficient frontier hyperbola, F represent the risk-free asset, and C is the combination of portfolios P and F.

1.7.2 Capital Asset Pricing Method

CAPM is a general equilibrium model that allows the measure of risk for any asset, as well as the relationship between risk and expected return for any assets in markets with equilibrium. CAPM is the first equilibrium model developed, and it is based on a set of inflexible assumptions, which makes it somehow objectionable for some experts.

CAPM Underlying Assumptions

The basic assumption of the CAMP model are introduced in [38] and [12]. Following is a brief listing of the adopted model assumptions:

- No transactions cost. This assumption may be true for some kinds of assets, but not all. If the model includes non financial assets, then, this assumption would be somehow critical.
- Assets are infinitely divisible. This may not be contradicting, but also for the literal case of financial assets only.
- No income taxes. This assumption (also evident) it might contradict with tax-exempt securities.

- Single agents cannot affect prices. In case of big pension funds, prices might be affected, in particular when unloading a big equity holding.
- Investors care only about mean and variance of their total financial portfolio or asset returns follow the normal distribution. This is evidently unrealistic for many investors.
- Unlimited short sales allowed. The individual investor can sell short any amount of any share.
- Unlimited lending and borrowing at the risk-less rate. This might be logical but some traders actually pay a spread between borrowing and lending.
- All investors have identical expectations. In other words, investors are assumed to be concerned with the mean and variance of returns.
- All investors have the same time horizon, this would well comply to average of people's expectations.
- All assets are marketable.

Standard CAPM

The CAPM has been developed with several forms that vary in rigor and mathematical complexity. We are going to address the standard CAPM which is a simple intuitive derivation of CAPM.

Following the assumptions of CAPM, concerning lending and borrowing, and having homogeneous expectations, leads to investors holding the same risky portfolio [38]. Therefore, in equilibrium it must be the market portfolio. The efficient frontier that people would face according to these conditions will be typically the one shown in figure 1.5. The straight line in the figure refers to the capital market line, where all investors portfolios and all efficient portfolios lie. The equation of the market line here is the same as that connecting a risk-less asset to a risky portfolio, estimated as follows:

$$\overline{R}_e = R_F + \frac{\overline{R}_M - R_F}{\sigma_M}\sigma_e \tag{1.7}$$

Where σ_e denotes an efficient portfolio, $\overline{R}_M - R_F/\sigma_M$ represents the market price of risk for all efficient portfolios. This equation describes the expected return for all assets and portfolio of assets in the economy.

1.7.3 Arbitrage Pricing Model

The Arbitrage Pricing Theory APT is another asset pricing theory, it is a one-period asset pricing model. It asserts that the expected return of an asset can be modeled as a linear function of multiple factors. It also assumes that avoiding arbitrage over statistic portfolio of assets leads to a linear relationship between return and its co-variance with macro-economic factors. Therefore, APT helps in figuring out which assets are miss-priced and accordingly avoid over-priced assets and trade under-priced assets. This is achieved by monitoring the rate of return in the model, through comparing the actual price to the model expected price at the end of the study period. APT was first introduced by Stephen Ross in 1976 as a substitute to the Capital Asset Pricing Model, in a way that they both speak of a relationship between asset return and co-variance, where co-variance as in CAPM and MPT represents the risk factor.

APT Standard Model

The APT assumes risky asset returns and tends to follow a factor intensity structure according to the following formula.

$$r_j = \alpha_j + \beta_{j1}F_1 + \beta_{j2}F_2 + \cdots + \beta_{jn}F_n + \epsilon_j \tag{1.8}$$

where α_j is an asset j related constant, F_k is a systematic factor, β_{jk} is the measure of sensitivity of the $j^t h$ asset towards loading factor k, and ϵ_j is the risky asset idiosyncratic random shock with mean zero. Note that the idiosyncratic random shock is considered non-correlated to assets and factors.

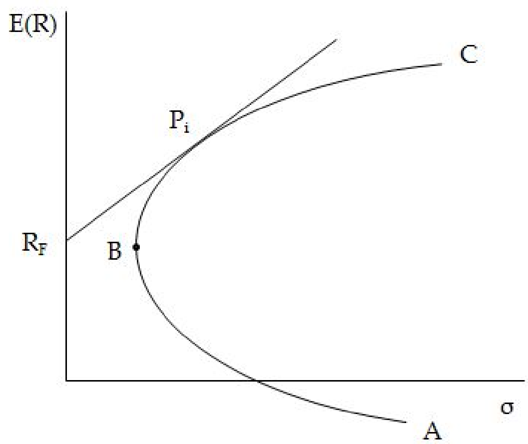


Figure 1.5: The Efficient Frontier with Lending and Borrowing

Another statement of APT is that the relationship between expected returns and factors becomes as follows (in case of assets following a factor structure):

$$E(r_j) = r_f + \beta_{j1} RP_1 + \beta_{j2} RP_2 + \cdots + \beta_{jn} RP_n \tag{1.9}$$

Where RP_k is the premium of the factor risk, r_f is the risk-free rate, assuming that the asset expected return is a function of its sensitivities to the n factors. For the validity of the estimation, two assumptions have to be considered, the market is perfectly competitive, and the number of factors never overcomes the number of assets.

CAPM Vs APT

APT and CAPM are two theories for asset pricing. The CAPM is based on its set of assumptions relating return and risk, while APT is less restrictive coming to assumptions giving more flexible model or asset return. Opposing the typical market portfolio, APT assumes that investors will hold a portfolio with specific arrays of β coefficients. One can think of CAPM as a specific case of APT where it is constrained to a single-factor model of asset pricing. Another way to describe the models with supply and demand, APT being a supply model and CAPM being a demand model. Since APT beta coefficients represent the sensitivity of the underlying asset to economic factors. CAPM is considered as a demand model, since its results are related to the investor's utility function.

1.8 Portfolio Evaluation

The most crucial part in any decision-making process is the evaluation part used for comparing the fund performance regarding other funds, and to verify how well did the fund follow the general policies. It is used by professional institutions and individuals, as well as personal investors helping them judge the performance and understand well the factors behind it, it can be thought of as results diagnosis. There are multiple portfolio evaluation techniques developed through time to judge the performance of a portfolio.

1.8.1 Sharpe's Rule

This rule was introduced by Williame F. Sharpe in 1994, and it measures the risk-adjusted performance of a portfolio. It is simply obtained by subtracting the rate of return for a portfolio from the risk-free rate, generally chosen as 10 year bond of the US treasury bills, Tbonds. Then, that measure is divided by the standard deviation of the portfolio returns. The formula of this ratio is as follows [61].

$$\frac{\overline{r}_p - r_f}{\sigma_p} \tag{1.10}$$

Where, $\overline{r}_p$ is the expected portfolio return, r_f is the risk free rate, and σ_p is the portfolio standard deviation. This ratio is used to asses whether the return of a portfolio is due to well modeled investment or is just a result of excess risk, since as it is already known the increase of risk a great factor for return increase. A high sharp ratio indicates a good risk-adjusted performance of the portfolio, while a negative ratio result indicates that the performance of the analyzed asset could be overcame by a risk-less asset.

1.8.2 Tranor Ratio

This ratio is also used for portfolio evaluation. It was first introduced by Jack Traynor. This ratio measures excess earned return, compared to risk-less investment per each unit of market risk. It is computed through subtracting the average risk-free return rate form the portfolio return, divided then by portfolio measure of volatility beta[21].

$$T = \frac{r_i - r_f}{\beta_i} \tag{1.11}$$

Where T represents Treynor ratio, r_i represents the return of portfolio i, r_f symbolizes the risk free rate, and β_i represents the volatility of portfolio i.

1.8.3 Jensen's Alpha

Jensen's alpha is another portfolio performance evaluation measure, that was developed by Michael Jensen in 1968 [35]. It is used to determine the abnormal return of a portfolio or security compared to the expected return theoretically. Normally, the theoretical return is predicted by the market model, usually using Capital Asset Pricing Model introduced in section 1.7.2. Calculating Jensen's Alpha is achieved through applying the following formula.

$$\alpha_j = R_i - [R_f + \beta_{iM} \cdot (R_M - R_f)] \tag{1.12}$$

Where α_j is the Jensen's Alpha measure, R_i is the portfolio return, R_f is the risk free rate, β_{iM} is the portfolio beta, and R_M is the market return.

1.8.4 Information Ratio

As the above described techniques Information ratio, or as also known appraisal ratio, is also a performance evaluation technique used for portfolio understanding and evaluation. It is a measure of the risk-adjusted return of a financial asset or portfolio of assets [17]. Normally, it is an estimate of the active return divided by the tracking error represented by the standard deviation of the active return.

$$IR = \frac{E[R_p - R_b]}{\sigma} = \frac{\alpha}{\omega} = \frac{E[R_p - R_b]}{\sqrt{\mathrm{var}[R_p - R_b]}} \tag{1.13}$$

Where R_p is the portfolio return, R_b is the benchmark return, $\alpha = E[R_p - R_b]$ is the expected value of the active return, and $\omega = \sigma$ is the standard deviation of the active return, or also called tracking error. The above ratio is a measurement to the active return of the manager's portfolio divided by the amount of risk that the manager takes compared to benchmark risk. The higher the resulting ratio, the better the portfolio manager is, and the other way around.

1.9 Conclusion

This chapter purposes in familiarizing the readers with the financial notions in this thesis and supply them with basic necessary understanding to financial investment and its various types. It clarifies the concept of financial markets and its security categories, and explains the trading basics from security analysis techniques to the concept of risk and return and the advantages of diversification. The second part gives a global look into portfolio management, its analysis selection and evaluation, and introduces the most used techniques for each. After referring to all details concerning the market and its securities, it becomes important to get in depth with the different available techniques of security analysis. Therefore, next chapter is dedicated for describing the two main types of security analysis, ie. fundemental and technical analysis and their deployed analysis tools.

CHAPTER 2 Fundamental and Technical Analysis, Introducing Technical Indicators

2.1 Introduction

The previous chapter widely introduced portfolio management and its analysis, selection, and evaluation steps. The security analysis as described earlier represents the first and most important step to building a portfolio. It is normally a subject of examining and studying in depth the strength of the security and the well being of its underlying company, to make sure whether deploying this security in the portfolio collection chosen enforces a future success in trading or lead to disappointment. As we mentioned earlier the security analysis is used to figure out the intrinsic value of securities. Efficiently analyzed securities in a portfolio can lead to highly profitable trades. Throughout time, multiple analysis techniques were developed where Fundamental and Technical analysis have proven to be the most efficient, which made them the resort of most traders. This chapter describes in depth these two main types of security analysis. It includes a detailed explanation of technical analysis, and its indicators where lies the core of this thesis motivation.

2.2 Fundamental Analysis

The first section of this chapter handles fundamental analysis. It includes an introduction to fundamental analysis, its qualitative and quantitative factors, an explanation of some major statements in fundamental analysis, its strengths and weaknesses, and finally a small conclusion to put things all together.

2.2.1 Introducing Fundamental analysis

It is the process of examining the underlying fundamental factors of a security that might affect directly or indirectly the status of the company, economy, or industry group. Similar to any security analysis technique, it aims to forecast future price movement and accordingly take advantage of the presumed change.

In fundamental analysis the technique differs with different security base levels. For example at the industry level, the analysis is directed towards the supply and demand of the product under study, while at the company level, the analysis becomes more concerned with the business plans, financial data, management strategies, and competition. As for the economic level, the analysis involves economic data that affect the growth or declination of the economy.

The basic supposition that this analysis technique lies on, is assuming that the market does not translate directly the value of a security. In other words, it supposes that the market takes some time to reflect the situation of a company on the price of its shares, giving analysts the ability to compute its actual worth before the market, and thus forecast the securities future price change.

The actual worth in financial terms is called the intrinsic value, to estimate this value fundamental analysts combine industry, economic and company analysis techniques. After estimating the intrinsic value of a security it becomes easy to judge whether this security is over or under valued, thus be able to foresee its future price movement. And as mentioned multiple times, knowing the future of a securities price allows traders to make a winning decision of buying or selling.

2.2.2 Quantitative and Qualitative Fundamental Factors

In fundamental analysis the factors that contribute in the analysis normally fall into two categories, the quantitative and qualitative factors. Similar to the regular definition of the two words, quantitative factors are the ones that are subject to numerical calculation normally anything concerning companies financial data, while qualitative factors are normally company aspects that are less tangible.

Quantitative Factors

Normally, each company has documents that are concerned with all its financial data. There are mainly measurable numeric characteristics of a certain business. Financial statements are considered to be the most important source of financial information of a company. Other than financial statements, there are many sources used to derive the quantitative factors. Examples of such sources are balance sheets, income statements, statements of cash flow, annual reports, footnotes, and much more. Fundamental analysts use these financial documents to determine certain measures and ratios for the quantitative part of the fundamental analysis [37]. Figure 2.1, lists all categorized quantitative and qualitative factors. Below is a brief introduction to each of the commonly used quantitative factors.

- **Earnings Per Share (EPS)**, usually earnings are used to indicate the expected dividends and growth potential of a company. The EPS ratio is calculated through dividing the net earnings of a company to the number of issued shares of a company. For example, if a company reports a year net earning of 20 million USD, and has 10 million outstanding shares, then the EPS of that company becomes 2 USD per share. This ratio plays an important role in comparing earnings of different companies.

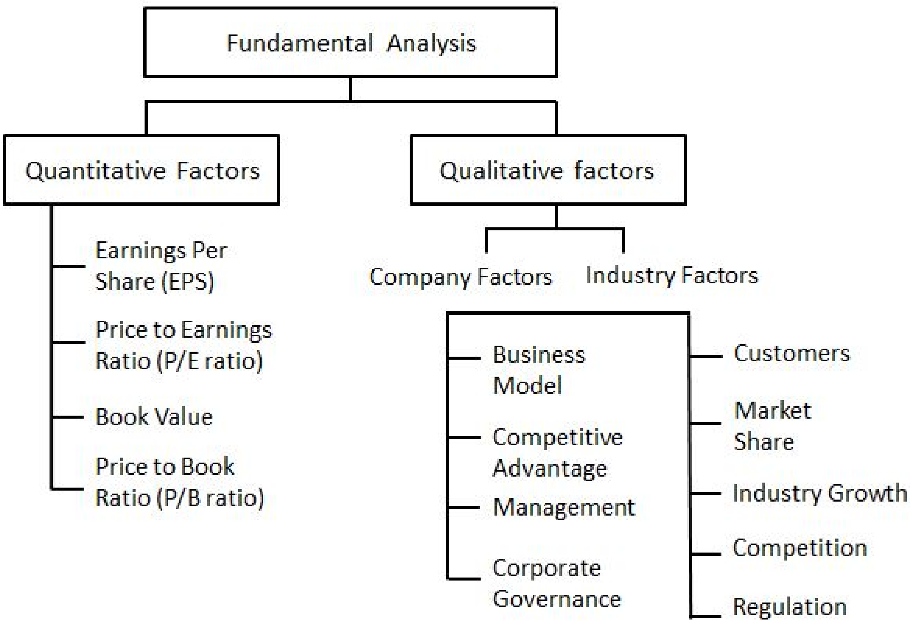


Figure 2.1: Schematic Diagram of Fundamental Factors

- **Price to Earnings Ratio (P/E)**, the P/E ratio is calculated by dividing the actual price of a companies share by its EPS. This ratio is used to study the truth of the ability of a company to pay its earnings. Generally a high P/E ratio indicates that the company is expensive, while a low P/E indicates that it is cheap.

- **Book Value**, this value measures how much of assets would a company have assuming a direct liquidation. It is calculated through measuring the total assets of a company taking out its liabilities. Comparing book values of various companies could be reached through dividing the latest book value by the number of outstanding shares of the company, this measure is known as the book value per share.

- **Price to Book Ratio (P/B)**, the P/B ratio is used to study the ability of the market to pay the company its hard shares. It is calculated by dividing the share price of a company by its book value per share, the higher the ratio is the more prepared the market is to pay above the hard assets of the company.

Qualitative Factors

Analyzing and assessing the qualitative assets of a company is considered relatively harder, since as mentioned earlier, they are less tangible aspects of the companies business. Nevertheless, studying qualitative factors is a very important part of fundamental analysis, where they play a very important role in estimating the intrinsic value of a company share. There are two categories of qualitative factors, the company and the industry factors. Check figure 2.1 for a clarifying categorization schema.

- **Company Factors**, These factors are related to the company and to the manner it uses in running the business. Although as already noted that quantifying the qualitative aspects of a company is difficult and sometimes impossible, they do have a great effect on the valuation of a share, and thus it is mandatory to consider them in the fundamental assessment.

  – ***Business Model***, this model can be simplified to be thought of as answer to two questions: what does the company do? and, how does the company make its money? This is normally understanding a business model of a company. In some cases, it is direct and easy to understand what a company does and what is its main way of getting money, but sometimes it gets complex to understand the model, but either way it is an important step to complete a full analysis of the company.

- ***Competitive Advantage***, this is another business consideration for analysts to examine. Normally, the ongoing or long term success of a company is greatly affected by its ability to maintain a competitive advantage. The competitive advantage of a company is its success in overcoming its rivals and sustaining this success.
- ***Management***, management is considered the most important qualitative factor of the analysis. Any business must have a leader to direct it towards success, and the performance of a business depends much on the steering of its management. It is not so easy to understand well the management of the business, since this is not something a company would make clear to the public. Nevertheless, there are some specific means open for investors to investigate and judge the management of a company like, conference calls, management discussions and analysis, past performance, and ownership and insider sales.
- ***Corporate Governance***, it can be explained as the ability of the company to protect its investors from illegal and unethical business activities. Analysts usually search to find out whether a company complies to the governance policies and regulations, and by that ensure whether the investors of this company are receiving there rights fairly.

• **Industry Factors**, these factors are examined by analysts to gain a deep understanding to the health of the financial company. Industry factors include all aspects of a companies functionality that would affect the market environment.

- ***Customers***, The number of customers that a company serves is related to the way its share value changes. A company that depends in its profit on a small number of customers, have more chance to lose profit since it depends on few sources, and vice versa.
- ***Market Shares***, having a clear idea about the size of a market share of a company helps in evaluating the strength of the company in the industry of which it belongs to. Companies with high percentage of shares in the market are considered among the strong market players, and this information can also indicate that the company has a strong competitive advantage that gives it a higher possibility to sustain its worth value.
- ***Industry Growth***, understanding the growth potential of the industry which the company belongs to, serves in having a better idea about the growth potential of the company itself. A market that is object to growth makes it easier for a business share to grow, since being in a market that has

less potential of including new customers makes it harder, since companies will have to gain customers from the ones already trading.

– ***Competition***, it is simply checking the number of competitors in the industry to understand the success potential of the company in competition. A company that operates in an industry with few alternatives, has a higher chance in gaining competitive powers over its rivals.

– ***Regulation***, although regulations play as good granting factors to the public, a highly regulated industry can be limiting to companies and this has a drastic effect on the companies ability to attract investors.

2.2.3 Information Evaluation

After collecting all the information about the underlying company and industry of the share, an investor is about to indulge in, the evaluation or interpretation of these information comes to place. The purpose of this evaluation is to determine and judge whether the company under study is considered of good value or not.

There are multiple ways for investors to measure or interpret there collected research information. One of the most commonly used methods is to compare its results with a peer group of companies. This enables the analyst to position or understand the placement of the company chosen among its vendors of the same industry, which definitely provides a precise indication of the company's relative value.

2.2.4 Strengths and Weaknesses of Fundamental Analysis

In this section, we briefly speak of the strength of fundamental analysis, along with its weakness and criticism.

Strengths

Although fundamental analysis is criticized by its complex detailed research needs, it does have some advantages. Here are some of the most common points of strength in this security analysis technique.

- **Long Term Trends**, one of the fundamental analysis strengths is its success with long term investments that depend on long term trends. It has its benefit in predicting the long term performance of a company on all trend levels, economic, demographic, technological, and consumer trends.

- **Value Spotting**, being able to evaluate a company from all its fundamental data is a strong tool to uncover the companies with valuable assets. Some of the most legendary investors known throughout history used the value spotting technique and made some inevitable revenue out of it.

- **Business Acumen**, this forms one of the most important advantages or points of strength to this analysis technique. Developing a complete understanding of the business on all levels gives the investors a deep precise understanding of its chances, sustainability, and lifetime. This knowledge can also help investors reveal and avoid companies that are subject to failure or shortfalls.

Weaknesses and Criticism

Despite the fact that fundamental analysis has its strength when correctly applied, it does come with some real constrains or disadvantages.

- **Time Constraints**: despite the fact that a sound fundamental analysis might result in good security evaluation, it can be very time consuming and exhausting mission. In some cases a fundamental evaluation can show an opposite or contradictory valuation to the one available in the market. This has caused skepticism among analysts and investors towards this kind of analysis.

- **Industry and Company Specific**: in fundamental analysis, there are various available techniques, specific for different types of companies. Using a certain technique might not always be the most efficient depending on the type of company under analysis. This makes a big confusion factor to analysts and can some times cause a whole research to be a waist of time.

- **Subjectivity**: since the valuation process in fundamental analysis depend on a group of assumptions generated from various time consuming studies, a slight change to the factors can cause a drastic effect to the judgment. This makes fundamental analysis highly sensitive, which drove analysts to develop different cases of valuation, which makes the technique even more subject to confusion.

- **Relative Efficiency Problems**: this is considered a big criticism to fundamental analysis, which comes from mainly two groups, supporters of efficient market hypothesis and proponents of technical analysis. The test of time is always considered to be the best judge of all judges. Ever since it came to existence, technical analysis became the major of security analysis followed by investors. The next section includes detailed description of this analysis technique. To briefly clarify the cause of its success over fundamental analysis, it uses some charts and tools that solely depend on the price and volume of the market, assuming that any fundamental data of a security is already included in its price giving no importance to all fundamentals of the company. Though, it is possible for both analysis to be used together, technical analysis assumptions cancel any value of fundamental valuation.

2.3 Technical analysis

Technical analysis is the attempt to forecast a security future price movement, through analyzing its historical data. Technical analysts believe that the future can thus be found in the past feature. They also assert that the fundamentals of security values are all summed up by its price. Therefore, they resort to using technical indicators to study patterns, trends, and some other price factors, and accordingly make their investment decisions [3]. Technical analysis had long been regarded with skepticism and doubt of its effectiveness. However, the accumulating evidence of market inefficiency caused a revival of academic interest in technical analysis claims. Since then, it has been showing great predictive powers compared to other strategies and analysis [48], [50] and [24]. In this section, we present the history of this security analysis technique explaining all its technical notions, charts trends and the indicators it indicators.

2.3.1 History

Technical analysis was an extension to the Dow Theory concept that was developed by Charles Dow around 1900, he was the editor of wall street journal[3].
Throughout his long journey in Wall Street and his everyday indulgence in the markets, he noticed that markets and stocks move in tandem. Whenever the market trends upwards, most stocks move along and vice versa. Therefore, for the purpose of interpreting the markets behavior, he developed two indicators the DJIA and the DJTA, Dow Jones Industrial Average and Dow Jones Transportation Average.
His proposed theory, the Daw Theory, was the root where technical analysis stemmed from. It introduced so many believes such as the principle of price summing up all information, the trending attribute of price, convergence and divergence, support and resistance, and many other principles. Thus, the endowment of Charles Dow to technical analysis cannot be undervalued.

2.3.2 Rational of Technical Analysis

The main assumption behind technical analysis is that price discounts all information of a security even the fundamentals of its underlying business. The other main belief is that past can be found in the future, or in a simpler notion, history repeats itself, thus one can foresee the future through analyzing the past performance of security prices.

It is agreed that the price of a security is the price that one agrees to buy and another agrees to sell. Then, this price depends much on the expectations of individuals integrated in this action. This forms a great deal and was subject to many debates and disapproval on the efficiency and rationality of markets and the effect of humans on its performance (see section 1.5.3), a later section includes a deep discussion on this matter. This undeniable human involvement is also the main reason behind the difficulty of consistent success in forecasting future price movement, regardless on the analysis technique or trading system used. The same reason made fundamental analysis very doubtful, since it can be very precise and takes into consideration all interfering fundamentals and still fail to meet the market real fluctuation,[4].

The main assumptions of technical analysis all together can be summarized as the following:

- Supply and demand are the factors that determine the security price change in the market. Any change in the supply and demand will be sooner or later translated into the market.
- Supply and demand are administered by various factors such as necessities, mood guesses, intuition, opinions, etc.
- Leaving out the minor changes in price, it is assumed that the long-term change of price follows a clear trend.
- Price chart patterns tend to repeat on certain circumstances, this recurring of prices can be a strong tool to predict future price changes.

2.3.3 Price Fields, Charts, and Patterns

Since technical analysis depends mostly on studying prices and their chart patterns, this section will explain the available fields, type of charts that are normally used to describe a security, and patterns that analysts seek to recognize in charts.

Price Fields

It has been mentioned earlier that one of the most important assumptions of Technical Analysis, is that price discounts all fundamental information of a security. Thus, the basic step for understanding this type of analysis necessitates a recognition of the different price fields used in the financial Market. Below is a list of different price fields available with a brief explanation to each field.

- **Open**, this is usually the price of the first trade for a certain time period. This price normally symbolizes the price agreement following the fight throughout the preceding period.
- **High**, this is the highest value a price reaches during the period studied. It represents the highest point where the number of sellers exceeds the number of buyers.
- **Low**, almost opposite to the high, it is the lowest price traded during the studied period, where it represents the lowest point through out the period, where buyers exceed the sellers.
- **Close**, as its name indicates it is the price that trades closed at towards the end of a period. Simply explained, it is the price of the last trade of a period. This is the most used price with analysis.
- **Volume**, it represents the number of shares that are traded during a period, or the volume of trading. The effect of volume is very important in indicating a begin or end of a trend, an increasing volume accompanying an increase in price forms an important indication.
- **Open Interest**, it is the number of all outstanding shares, or contracts that are available. This is often used as an indicator itself.
- **Bid**, this is the price the market maker is willing to give for buying the security.
- **Ask**, contrarily this is the price the market maker is willing to take for selling the security.

Charts Types

The earlier discussed price fields form the basis of all developed indicators out there, hundreds of technical tools are developed and every single one uses one or more of the above fields for its estimation. All the above fields are plotted with different charting techniques to help analysts and traders to recognize changes and patterns easily. Explained simply, charts are graphical representations of price fields over a set time frame. The major used chart types are described below [22].

- **Line Charts**, are considered the most basic type of charts, it is mainly a linear plot connecting close prices of a certain security over time. The used price field for these charts is normally the close price, since it is often considered as the most important among other price fields (high, low, open). Figure 2.2 introduces an example of a typical line chart.

- **Bar Charts**, the bar chart has more details in it than just the close price used for line charts, it illustrates the open, close prices of a security, over a time frame. It is constructed of consecutive vertical lines with two horizontal dashes on each. The horizontal dashes are one to the left representing the open, and another to the right representing the close. Whenever the open is lower than the close, the line is plotted in red indicating an increased value of the security at the period. On the other hand, a higher open indicates a decreased value to the security, and is usually represented by a black colored line. Refer to Figure 2.3 for clarification.

- **Candle stick Charts**, it is very similar to bar charts with the difference of using bars instead of horizontal dashes to represent the high and low price values, distinguished by the width of each bar. The main confusion about this type of charts is that it does not have an agreed coloring scheme as in bar charts, where different sources use different coloring to indicate upwards and downwards change of security values. Figure 2.4 includes an illustrated explanation of the candle stick chart.

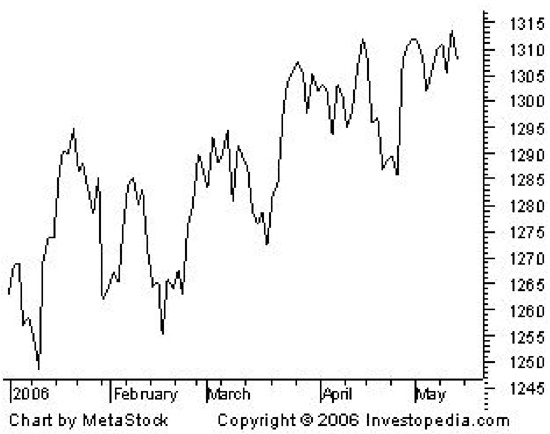


Figure 2.2: Line Chart Example

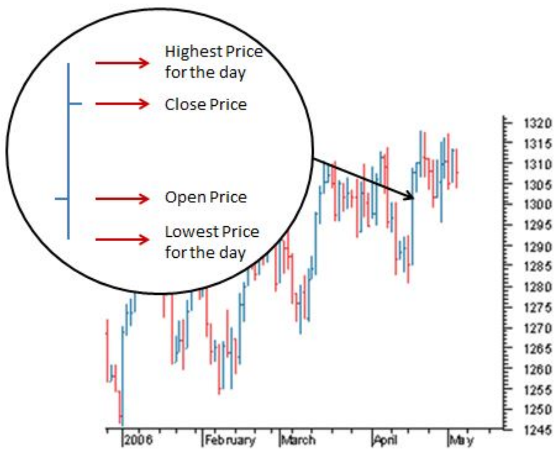


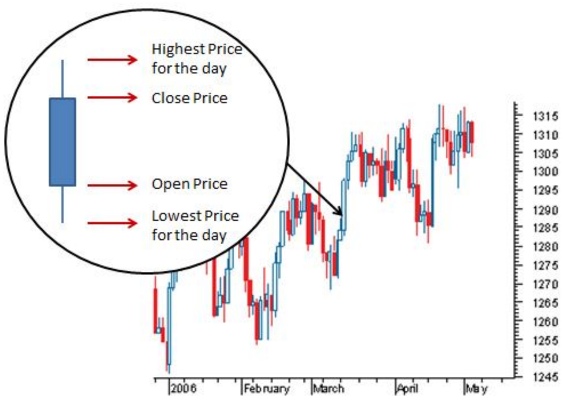
Figure 2.3: Bar Chart Illustration

- **Point and Figure Charts**, this type of charts is not widely used among investors, yet it was used when first technical analysis emerged. It is a plot of Xes and Oes, were Xes indicate an upward trend in a security price, and Oes indicate a downward trend. Analysts regard the point and figure chart as a way to eliminate price noise, and make it easier to spot trend changes with less confusion. Figure 2.5 introduces an example to this type of charts.

Chart Patterns

After understanding the essentials of Charts and used price fields, this section addresses the signals or patterns that analysts track in charts, to help them predict future movements of security prices. The theory behind seeking patterns in technical analysis comes from the assumption that future repeats itself. Thus, throughout time analysts developed a series of patterns that has been noted to precede certain price movements. The goal of pattern recognition in technical analysis is to identify trading opportunities before they happen. Although, these patters have shown success various times over and over, its results are not always granted and depend much on the trader's manner of the analysis. For that purpose, many would consider chart pattern interpretation an art rather than a science. Below is a list of the most used chart patterns by technical analysts.

- **Support and Resistance**, these are the levels where price fights to cross upwards and downwards (refer to figure 2.6 for demonstration). The upper level is called the level of resistance, and the lower level is that of support. There analogy goes to the ongoing battle between buyers and sellers in the market, where the buyers push prices higher and sellers push them lower. The resistance for example is formed when buyers try to take control of prices and sellers resist to prevent them



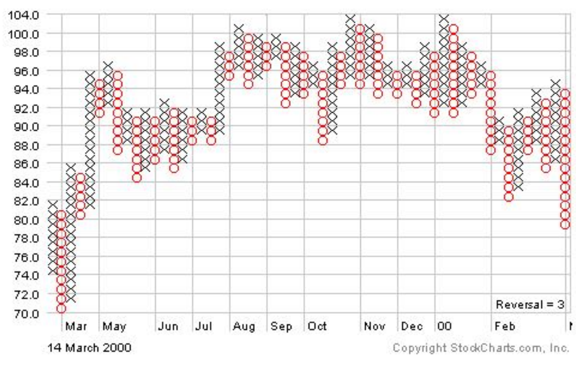


Figure 2.5: Point and Figure Chart Exam-

Figure 2.4: Candle Stick Chart Illustrationple

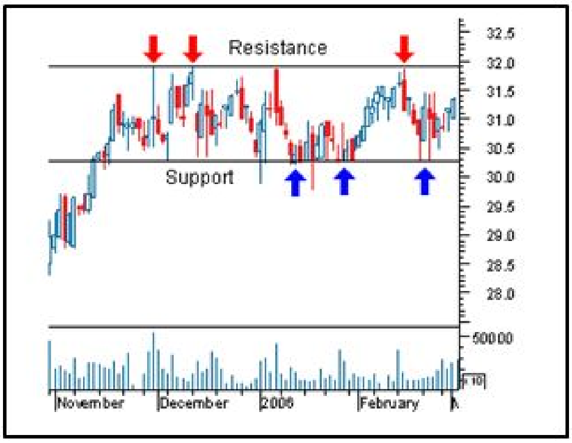


Figure 2.6: Support and Resistance Levels

from going higher than a certain price. The support is the opposite where sellers are able to take control and support price, preventing it from decreasing further than a certain level.

- **Head and Shoulders**, there are two types of this pattern: the normal head and shoulders and the reverse one. The main indication of this pattern is the trend reversion. When a security is witnessing an up-going trend and a head and shoulders patterns occurs, then it is relieving a weakness in the trend. It is considered as a signal of trend reversal, meaning that the up-going trend is about to reverse into a downtrend. The inverse head and shoulders pattern is the opposite, whenever it occurs during a down trend it is considered as a coming reversal of a trend upwards. Figure 2.7 includes a graphical demonstration of the head and shoulders patterns.

- **Cup and Handle**, This trend is normally a confirmation of a bullish (upward) trend. It is a cup like pattern where an upward trend forms a gradual descending and reascending, followed by an upward sideways fluctuation forming the handle. The trend continuation is confirmed when the resistance level formed in the handle is crossed upwards. Figure 2.8 shows a cup and handle pattern formation.

- **Double Tops and Bottoms**, This is among the most common and reliable patterns analysts resort to. It is also a trend reversal indicating pattern. It occurs in an uptrend when the price forms two highs trying to cross a certain resistance levels with no success, and is thus considered a sign of trend reversal. In the case of a downwards trend, bottoms form where the price finds support and does not cross it down, it is considered as an upward reversal signal. Figure 2.9 exhibits both tops and bottoms patterns.

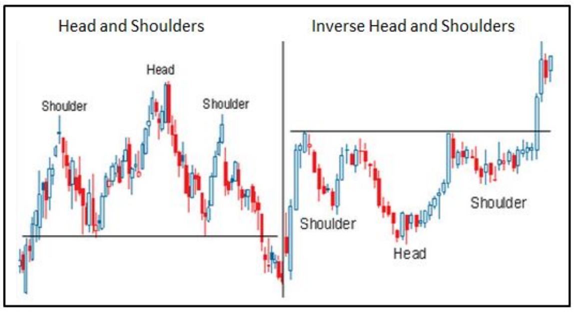


Figure 2.7: Normal and Inverse Head and Shoulders Patterns

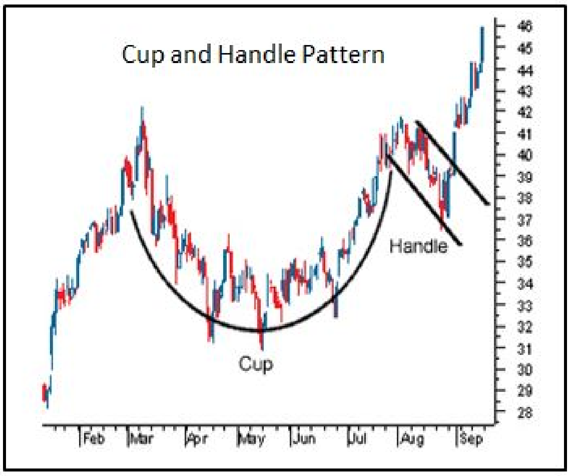


Figure 2.8: Cup and Handle Patterns

- **Triangles**, this is also another pattern commonly used by analysts that has three types: symmetrical, descending, and ascending triangles. As illustrated in figure 2.10, the symmetrical triangle is where two trend lines converge towards each other, and whenever a cross to either trend lines happen is considered a trend continuation in that direction. The ascending triangle has a flat upper trend line, and an bottom trend line inclined upwards, while the descending triangle is formed

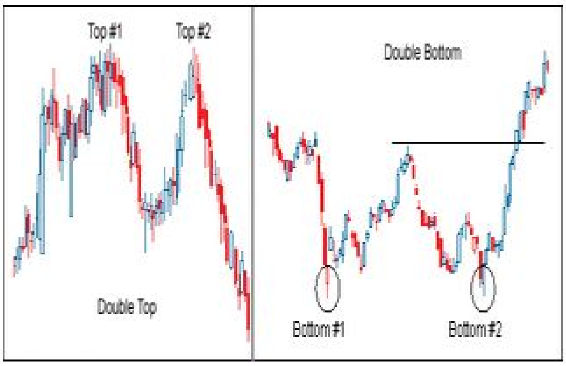


Figure 2.9: Tops and Bottoms Patterns

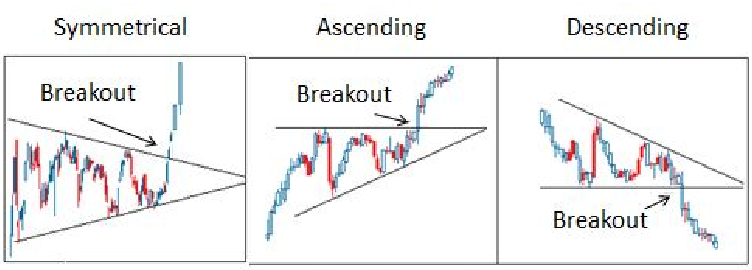


Figure 2.10: Symetrical, Ascending, and Descending Triangle Patterns

in an opposite direction. The purpose of the ascending and descending triangles is tracking upward or downward breakouts.

- **Flag and Pennant**, these patterns are both confirmation patterns used to confirm the continuation of a trend. The time where they happen is usually through sideways price movement and is complete when the price continues its propagation in the same trend direction. Figure 2.11 demonstrates both patters real time charts.
- **Wedge**, this type of patterns is similar to the symmetric triangles pattern, with the difference of being oriented either upwards or downwards. It is a signal of trend confirmation or reversal, which makes this pattern analysis confusing to some analysts. But the main assumption is that an upwards oriented wedge is considered as an indication of a bearish trend, and downwards wedge is a signal of a bullish trend (Figure 2.12).
- **Gaps**, as its name indicates, a gap is an empty space in the price chart. It is normally formed by a sudden significant change in price indicating an important incident concerning the security. Candle stick and bar charts show gaps when they occur, while line charts do not.
- **Triple Tops and Bottoms**, they are also trend reversal signals almost identical in functionality to double tops and bottoms introduced earlier. However, they are less often prevalent. An illustrative example could be found in figure 2.13.
- **Rounding Bottom**, this is also a trend reversal indicating pattern, that is very similar to the cup and handle, but without the handle form. The lack of the

Figure 2.11: Flag and Pennant Patterns

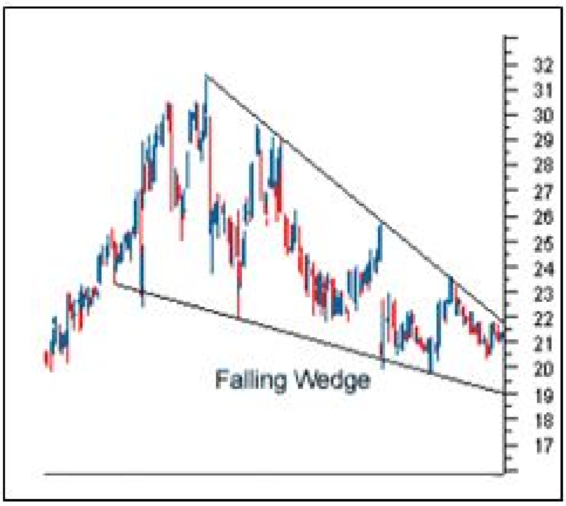


Figure 2.12: Wedge Patterns

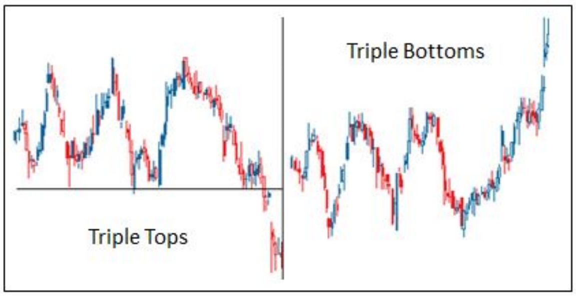


Figure 2.13: Triple Tops and Bottoms Patterns

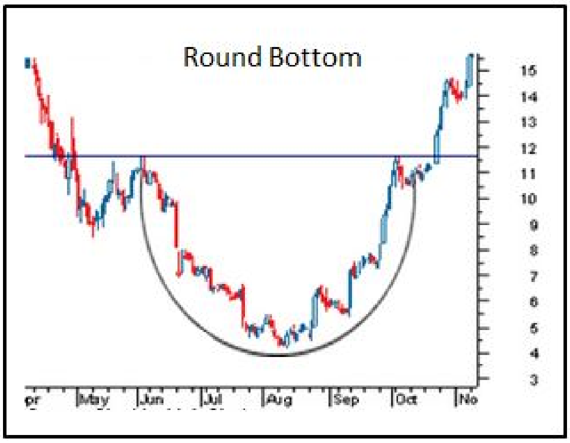


Figure 2.14: Round Bottom Patterns

confirmation that the handle plays, makes this pattern recognition and analysis quite confusing to traders. An illustrative example can be found in figure 2.14

2.4 Technical Indicators

Technical Indicators are simple mathematical formulas applied to the price and volume of securities to confirm future movement of security prices and make accordingly buy or sell actions. Along to the price data, they add information to help traders understand well and analyze factors like money flow, trends, momentum, and volatility. Indicators fall into two categories: the leading and the lagging indicators. Leading indicators are the ones that lead price and are used for price forecasting purposes, while lagging indicators lag behind the price which gives them the strength of confirming or denying current expectations. As for the construction of indicators, it can be distributed into bounded, and non-bounded indicators. Bounded Indicators are those that normally fall between two levels, the overbought and oversold levels, they are often known as oscillators and they are the most common construction of developed indicators.

2.4.1 Crossovers, Divergences, and Breakthroughs

The way analysts follow indicators to generate buy and sell signals is through inspecting breakthroughs, crossovers and convergence divergence. Below is an overview of each, and a graphical demonstration on such occurrence in real time charts.

- **Crossovers**, are considered as the most popular, and they usually occur when price crosses over a certain signal line like moving averages, or even when a certain moving average crosses another. As an example of how it is used to indicate buy and sell signals, we address the center line cross over. It is mainly applied on oscillators that fluctuate above and below a center line. A bullish (buying) signal is generated when the oscillating indicator crosses above the center line. A bearish (sell) signal is generated when the oscillating indicator crosses below the center line.

- **Divergences** , it is another way to use indicators for deriving buy and sell signals. It takes place when the price trend and the indicator trend contradict in direction, which indicates a weakness in the current trend and a possibility of future reversal.
  - ***Bullish Divergence***, is formed when the price of a security generates a lower low, while the Indicator generates a higher low. This indicates underlying strength in the security, thus would then be considered as a bullish or buying signal.
  - ***Bearish Divergence***, is formed when the price of a security generates a higher high, while the Indicator generates a lower high. This indicates underlying weakness in the security, thus it would then be considered as a bearish or selling signal.
- **Breakout** a Breakout takes place when a security price crosses a conceptual level, not a real signal as that of an indicator or moving average. By virtual level, we mean support and resistance levels or overbought/oversold levels, crossing these levels can reveal important occurrences that are not always shown by normal crossovers.
  - ***Support Resistance Breakouts***,
    * *Resistance breakout*, it occurs when the security price or indicator breaks up through its level of resistance, this is considered as a bullish (buying) signal. Sometimes the broken resistance level becomes the new support. This type of breakout comes usually after a bullish divergence and works as a confirmation signal.
    * *Support breakout*, it occurs when the security price or indicator breaks down through its level of support, this is considered as a bearish (selling) signal. Sometimes, the broken support level becomes the new resistance. This type of breakout comes usually after a bearish divergence and works as a confirmation signal.
  - ***Overbought Oversold Breakouts***, as mentioned earlier, some technical indicators are bound oscillators; they are used to identify overbought and

oversold thresholds. When the indicator crosses the overbought threshold, a bearish signal is considered. Whereas, when the indicator crosses the oversold threshold, a bullish signal is considered.

2.4.2 Technical Indicators

This section covers an overview about the concept of an indicator, it is simply a series of data points derived by applying simple mathematical formulas to past prices or volume data of a security. Furthermore, the different categories of indicators and their pattern recognition are altered earlier in this chapter, where in this section a list of the 10 most used indicators that passed the test of time and proved to be highly performing among hundreds of developed indicators throughout history are introduced correspondingly. The following indicators of different types are to be used later in the process of testing and evaluation of the decision support systems.

* **Relative Strength Index (RSI)** This is a momentum indicator that was developed by J. Welles Wilder [68]. It was introduced in his book "New Concepts in Technical Trading Systems" in 1978, since then it became extremely popular, were it attracted the attention of many traders and was featured by various books and articles. It oscillates between the values of 0 and 100 with indicated overbought and oversold levels of 70 and 30 correspondingly. Whenever RSI crosses upwards the level of 70, the studied security is considered overbought, on the contrary when a downwards cross of RSI to the level 30 occurs, it is counted as an indication of the security being oversold, refer to figure 2.15. Analysts also

Figure 2.15: Relative Strength Index Overbaught-Oversold Leves

observe crossovers and divergences for signal generation.
Mathematical Representation:

$$RSI = 100 - \frac{100}{1 + RS} \tag{2.1}$$

$$RS = \frac{Average\ Gain}{Average\ Loss} \tag{2.2}$$

Where the average gain and loss are calculated over a specified period of time, with the default period recommended by Wilder in his book being 14-days [68].
RSI has preserved its place among traders frequently used indicators, and has well passed the test of time. Despite the changes that the market has witnessed since the day RSI was developed, it is still chosen for its relevancy to help analyze the market.

* **Simple Moving Average (SMA)** This is a "trend following" indicator used to define the current direction of a trend through smoothing the price. Moving averages are trend lagging indicators since there calculations are based on past prices, thus they are used for confirmation rather than forecasting future trend movement. They can be though of as indicators to filter the noise from the price data, hence facilitate price analysis. As for the simple moving average particularly, it is a non weighted moving average that estimates the average of the closing price of a security over a certain period of time, while continuously discarding old data and include new ones as they come to availability. Thus, the first SMA is the normal average of prices over the period chosen, as it moves to the next day it drops the oldest date price and adds the new date price to its average calculation maintaining the data points to the period specified.
  Mathematical Representation:

$$SMA_d = \frac{\sum_{i=1}^{n} P_{(d-i)+1}}{n}, \quad n \leq d \tag{2.3}$$

  Where SMA_d is the simple moving average at day d, n is the number of days chosen to be the period used, and P represents the closing price at a certain day. Figure 2.16 illustrates two SMAs with different periods witnessing a crossover.

* **Exponential Moving Average (EMA)** This is also a trend lagging indicator, belonging to the same family of SMA. The added value this indicator offers over SMA is that it reduces the lagging by giving more weight to the most recent price entries. The period chosen usually for

Figure 2.16: Two Simple Moving Average Cross-Over

Figure 2.17: Exponential Moving Average Illustration

moving averages is related to the lag needed by the analyst for the monitoring process. For example a long period moving average is used for long-term trend confirmation and price smoothing. Hence the usage of moving average and their selected period depends on the objective of each analyst.

Mathematical Representation:

$$EMA_d = P_d - EMA_{(d-1)} \times multiplier + EMA_{(d-1)}. \tag{2.4}$$

Where EMA_d represents the exponential moving average at day d, P_d is the close price at that day, and the multiplier is a constant value estimated by $2/(n+1)$ with n being the selected time period. Moving averages are sometimes used for trend identification, which is normally generated by crossovers of the moving averages to the price or the crossover between moving averages of different time periods. Figure 2.17 demonstrates an example of EMA showing the patterns of support in an uptrend and a resistance in a down trend.

∗ **Moving Average Convergence Divergence (MACD)** MACD is both trend following and momentum indicator that studies the relationship between two moving averages. The involved moving averages are the MACD, which is a difference between 12-day exponential moving average (EMA_{12}) and the 26-day exponential moving average (EMA_{26}) and the signal line which is an exponential moving average of the $MACD$ signal itself.

Mathematical Representation:

$$MACD = EMA_{12days} - EMA_{26days} \tag{2.5}$$

$$Signal\ Line = EMA(MACD)_{9days} \tag{2.6}$$

Bullish (buying) signals are generated when the MACD signal crosses the Signal line upwards, while bearish (selling) signals are generated when the MACD signal crosses the Signal line downwards (2.18).

∗ **Commodity Channel Index (CCI)** CCI was originally developed by Donald Lambert and introduced in "Commodities magazine" in 1980 [44]. It aids traders in identifying cyclic patterns of securities and recognizing the occurrence and reversals of trends. It is a typical momentum oscillator type of indicators that fluctuates between the levels of -100 and 100, enabling analysts to identify when the asset is overbought or oversold. A cross above the 100 level asserts that the security is being overbought generating a selling signal, while a cross below the -100 levels asserts that it is being oversold hence generating a buying signal. Refer to Figure 2.19 for a clear illustration.
Mathematical Representation:

$$CCI = \frac{TP - MATP}{c.MD} \tag{2.7}$$

Where TP is the Typical Price which is the daily average of the high, low and closing prices of a security; $MATP$ is the moving average of TP over N-period of time; MD is the mean deviation which is the average difference between TP and $MATP$, and c is a constant with a default value of 0.015
The extremes used by oscillator momentum indicators can be subject to change in certain cases, depending on the volatility of securities, where relatively volatile securities may require farther extremities than docile ones.

Figure 2.18: Moving Average Convergence Divergence Illustration

* **Bollinger Bands (BB)** This is a volatility type indicator developed by John Bollinger in the 1980 [9]. It mainly includes three bands, the upper, middle, and lower bands (Figure 2.20). The outer bands are normally allocated two standard deviations above and below the middle band. These bands almost act as moving average envelopes of the price. Bullish signals are generated when the Price signal crosses above the upper band, middle band, or lower band while bearish signals are generated when the price signal crosses below the upper band, middle band, or lower band.
  Mathematical Representation:

$$MiddleBand = SMA_{20days} \tag{2.8}$$

$$UpperBand = SMA_{20days} + c\alpha \tag{2.9}$$

$$LowerBand = SMA_{20days} - c\alpha \tag{2.10}$$

  Where SMA is a simple moving average over a period of time, c is a constant with a default value of 2, and α is the 20-day standard deviation of price.

* **On Balance Volume (OBV)** This a volume based indicator was introduced by Joe Granville in his 1963 book "Granville's New Key to Stock Market Profits" [34]. It measures the pressure of trading in the market, in a culmulative manner through adding the volume on its rising times and subtracting volume on its falling times. OBV can be applied on securities following the change of close prices, or it can be applied to the market as a whole. Figure 2.21 shows OBV while it reveals a bullish divergence.

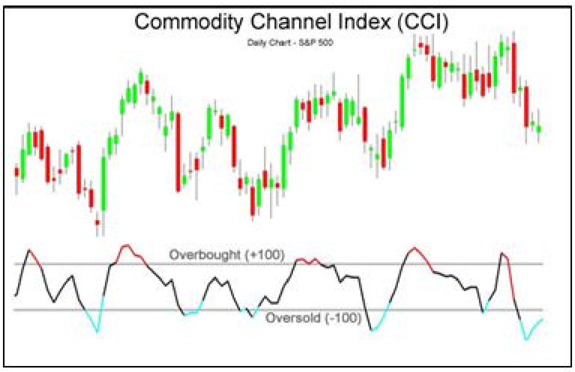


Figure 2.19: Commodity Channel Index Overbaught-Oversold

Figure 2.20: Bollinger Bands Illustration

Mathematical Representation:

$$OBV_d = OBV_{d-1} + \begin{cases} Volume & if\ P_d < P_{d-1} \\ 0 & if\ P_t = P_{d-1} \\ -Volume & if\ P_d > P_{d-1} \end{cases} \tag{2.11}$$

Where OBV_d is the on balance volume on day d, OBV_d is that of the previous day, P_d and $Pd-1$ is the closing price at the current and previous day correspondingly.

The change of price, increasing or decreasing controls whether the volume is assigned negative or positive. In the case where the current closing price is higher than the previous one, the volume gets a positive value. On the contrary if the current price is less than the previous one, then, the volume gets signed negatively. Therefore, the move of OBV and the price is positively proportional. To understand more the way of analyzing this indicator, lets assume the price generated a high at a certain time, accordingly OBV will similarly generate a high. If price makes a new higher high, and OBV fails to overcome its previous high, then, this is considered as a negative divergence, which mean a prediction of trend weakening or a selling signal. A positive is the exact contrary case where OBV fails to generate a lower low while price succeeds in doing that, generating a strengthened trend future, or a buying signal.

* **Rate of Change (ROC)** This indicator is a momentum oscillator, that studies the speed at which prices change over time periods through com-

Figure 2.21: On Balance Volume Revealing Bullish Divergence

paring the current security price with the price a period of time ago. Being an oscillating indicator ROC fluctuates above and below the zero level from positive to negative and vice versa (Figure 2.22), measuring the rise and fall of price throughout time. The patterns monitored with ROC indicator are normally crossover the zero line, divergences, and overbought-oversold examination.
Mathematical Representation:

$$ROC = \frac{P_d - P_{d-n}}{P_{d-n}} \times 100 \tag{2.12}$$

Where P_d is the closing price at day d, P_{d-n} is the closing price n days ago. Therefore an upward thrust of ROC represents an advance in price, while a downward thrust of ROC implies a decline in price. A sustained positive or negative reading in ROC can be used for trend confirmation. It is a momentum indicator that oscillates between 0 and -100 (Figure 2.23). Its main purpose is comparing the close price to the high-low range of a certain time period. The signals generated by this indicator come from analyzing crossovers, and monitoring overbought and over sold levels. Since W%R oscillates between 0 and -100, the level -50 is considered to be the center line. Crossing above the center-line indicates that prices are trading in the upper half of the high-low range, while a cross below that level indicates trading in the lower half of the studied period high-low range.

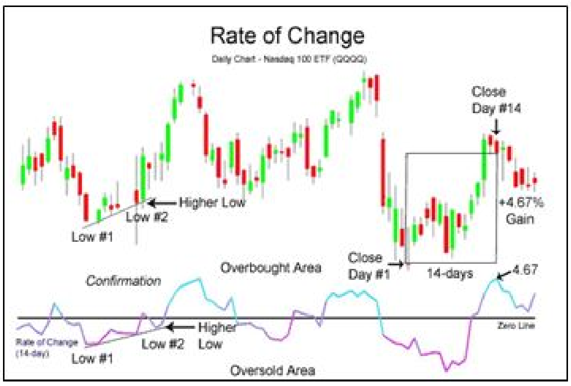


Figure 2.22: Illustrative Image for the Rate of Change

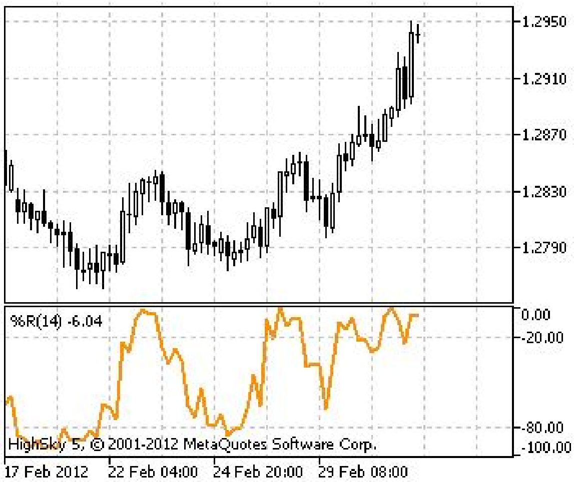


Figure 2.23: Illustrative Image for William % Rule

Mathematical Representation:

$$\%R = \frac{(HH - P)}{HH - LL} \times -100 \tag{2.13}$$

Where HH is the highest high along the studied period, P is the closing price, and LL is the lowest low along the studied period. The multiplication by -100 is used to correct the inversion. Overbought and oversold levels are also monitored with W%R to generate buying or selling signals, where the level -20 represent the overbought threshold and -80 represents the oversold threshold. A cross above the -20 level generates a selling signal, and a cross below the -80 level generates a buying signal. A strong indication is considered, when both crossover of center-line and either overbought-oversold crossovers confirm.

* **Linear Regression Indicator(LRI)** The LRI is a trend indicator that is used to determine the direction of trends. It is represented by a straight line that best fits the price between an ending and a starting point. Analysts consider this line as the fair value for price, where any deviation of price from this line will after all lead to a return to the linear regression line. Trading signals are generated by studying the crossover of prices to the linear regression line. A cross above the line represents a buying signal, where a cross below is considered a selling signal (Figure 2.24).
Mathematical Representation:

$$y = a + bx \tag{2.14}$$

$$a = \sum y - b \sum x \tag{2.15}$$

$$b = \frac{n \sum (xy) - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2} \tag{2.16}$$

Figure 2.24: Linear Regression Indicator Generating Buy and Sell Signals

Where x represents the current time period, and n is the number of periods used. Another popular way of deploying LRI is through constructing linear regression channel lines, which were first introduced by Gilbert Raff. The Channel is modeled by three lines the LR line with two parallel equidistant lines above and below. the distance of the two lines from the LR line is estimated through measuring the distance of the furthest close price from the LR line. These high channel is treated as a resistance level, and the lower channel is considered the support level, which are then used for buying and selling signal generation.

2.5 Conclusion

This chapter covered the principles of security analysis techniques and the controversy around fundamental analysis, and the effect of behavioral finance in depreciating the assumptions of fundamental analysis. This market irrationality theory contributed in spreading the interest in technical analysis among traders. The work motivation of this thesis is oriented towards technical analysis; in particular its technical indicator tools. The next chapter addresses the history of different using reasoning methods and artificial intelligence techniques in the filed of finance. It details the primary step in the work motivation of this thesis. It is a general pre-processing general approach which is based on Hybrid probability-possibility system that handles historical price data of different technical indicators and get it ready for the later steps of fusion techniques.

CHAPTER 3

History of Artificial Intelligence Technologies with Finance: The General Pre-processing Approach

3.1 Introduction

After covering all the needed knowledge on security analysis techniques, the controversy around fundamental knowledge, and the explained drift of trader interests towards technical analysis techniques, it becomes convenient to explain the path and orientation of the situation challenges and problems. This chapter gives an accurate introduction to the usual theories and assumptions logically explained to help the reader intuitively and clearly understand the reasons behind the proposed solutions and approaches detailed in the following chapters. Then chapter also puts forward a detailed description on the history of reasoning methods and their integration with finance along different aspects. This chapter forms simply the reflection of the challenges and problems of the environment under study and its tools and conditions.

3.2 A Logical Reflection

The previous chapters introduced readers to the objective of traders and analysts in predicting future price changes to grantee making revenue with discounted risk. The devotion to technical analysis as an effect to the growing controversy and skepticism surrounding fundamental analysis, in application with the arrival of behavioral finance to prove that markets are not all the time efficient and have human emotions interfering deeply in the market behavior, have also been explained. Therefore, it has became evident the need of a system to support traders in the process of taking the right decision of buying, holding, or selling at the right time.

The integration of human factors with financial markets and the uncertainty that is accompanied with technical analysis due to its dependence on many parameters, such as the change of indicator efficiency when applied on a certain stock or on different time horizon of the same stock, the way indicator signals are interpreted, and securities are chosen, necessitates the use of reasoning methods that can best handle such a situation to mimic and overcome the interpretation methods of typical analysis techniques.

The first challenge put into defiance is studying whether the integration of multiple indicators and merging their effect in a decision support system would be more profitable than following an individual indicator. The second challenge was choosing the right reasoning methods to deal with such conditions and perform the fusion and decision making process. Lets take a look on the aroused challenging problems with financial markets and technical analysis and the contribution of research on dealing with such challenges.

3.3 History of Reasoning Methods and Artificial Intelligence Technologies with Finance

A good starting point would be taking a look on previous applied methods, where various techniques have been applied for predicting market and security price movements, and portfolio risk evaluation. Multiple techniques from neural networks, to fuzzy logic, probability, genetic algorithms, machine learning and pattern recognition, have been integrated to finance in order to reach the target of achieving maximum profit with minimum loss.

3.3.1 Visual Technical Pattern Recognition Approaches in Finance

One of the known challenging subjects in technical analysis is detecting visual technical patterns that closely mimics the recognition of a human expert. Many approaches where proposed to defy challenges integrated with financial technical analysis such as, the inclusion of cognitive uncertainty and the manner of correctly accommodating it with pattern recognition.

For instance, Zhou and Dong [70] incorporate the cognitive uncertainty of technical analysis by using a Fuzzy logic-based system. Their approach confronted the ability to precisely detect and interpret technical patterns, compared to usual visual pattern analysis techniques applied by experts. Their proposed approach uses fuzzy logic, making use of Zadeh's assertions of the feasibility of introducing the cognitive uncertainty into the process of automatic detection. The latter proposed system makes use of a Gaussian kernel-Based Smoothing to capture price data in an accurate manner, avoiding the effect of noise on the detection process. The smoothing they propose, is then followed by an automating with a sequence of five consecutive local extremes, forming a pattern template to follow. Then, Fuzzification is applied to the pattern templates (see section 2.3.3). Finally a testing and evaluation process is included using Cumulative Abnormal Returns (CARs) as a measure for performance comparison and evaluation applied with different parameters.

Others have examined different aspects of visual technical patterns such as, Leigh, Purvis, and Ragusa [45]. They have proposed a decision support system that combines the methods of technical analysis, pattern recognizer, neural networks, and genetic algorithms to forecast the NYSE composite index. Also, Levy in [46] studied the predictive significance of the Five-Point chart patters applying tests on 32 possible forms of this pattern. Testing was applied on historical data of the NYSE, with Rate of Return being the measure of evaluation along to its standard deviation from the market. Levy marked a contradiction in decision signals of different forms of the same pattern type (the Five-point type), where neither of the pattern forms performed differently from the market. This study stated that in the studied stocks (US stocks) no predictive power of patterns was noted. Another study concerning patterns was introduced by Brock, Lakonishok and LeBaron [11] that used the Dow Jones Index historical data to examine the moving average and trading range breakouts, and accordingly developed a trading strategy that follows buy and sell signals following these levels breakouts. The results showed success return wise and were considered informative.

Much more studies concerning pattern recognition in technical analysis, and also well reputed consultancy companies have been deploying the successful research techniques of visual technical pattern recognition, to facilitate the decision making process for traders. Furthermore, technical pattern recognition helps analysts and traders take a wider look on the situation, helping avoid the distraction of non informative price data noise, helping grant return with less risk of error. Moreover, with machine learning and pattern recognition, it becomes easier for averagely experienced traders to capture slight differences that might have an important effect on the decision making. On the other hand, there are some limitation when resorting to visual technical pattern recognition. For example the exact imitation of human reasoning taking into consideration intuition is not straightforward, and is not often accurately achieved losing the ability of human interpretation. In addition, the definition of some patterns differs from expert to expert, which causes a controversy when trying to mirror experts definition technique and interpretation.

It is fairly decided that pattern recognition have shown success in its deployment. Where, many financial consultancy institutions follow and adopt its evolution to give precise aid to its customers. The uncertainty with the financial market is beyond patterns. It is correlated with security selection, indicators efficiency and parameter selection, market efficiency and rationality of traders, the decision making process itself and many other aspects not to be neglected as well.

3.3.2 History of Fuzzy Systems, Genetic Algorithms, and Trading Rules with Finance

Although predictability of the financial market is always the subject of ongoing debates, one cannot ignore the successful research progress of many reasoning methods and artificial intelligence techniques when integrated with technical analysis. Many different studies and innovated strategies have shown great achievements. Either with respect to generating better return or less risk than typical decision based strategies. As mentioned earlier uncertainty, ambiguity, and vagueness are all integrated with technical analysis throughout many different aspects. Starting by analyzing visual technical patterns, experts knowledge processing, the uncertainty accompanied with the forecasting and decision making, portfolio diversification and selection, ending with the inevitable human effect on market efficiency and rationality. The previous section 3.3.1 discussed the research and proposed strategies to handle the uncertainty with visual technical pattern recognition aspects, where we specified the challenges and contributions and progress of that specific research direction.

In this section, we address the history of using fuzzy logic and genetic algorithm for stock selection and the deployment of technical trading rules. In the addition to the fuzzy-based systems for visual technical pattern recognition mentioned earlier, fuzzy Logic have been used variously in dealing with some of the mentioned challenges in financial technical analysis. For example, Hiemstra [36] presents a stock market prediction approach and introduces in it the architecture of a fuzzy logic-based support system. Heimstra states that fuzzy logic is more preferable to be used with predicting financial market movements, in particular the case where the forecast is related to the experts manners of analysis. The paper first introduces a general scheme to predict the stock market. Then, it presents a fuzzy logic model based on the earlier proposed general scheme, followed by an evaluation of the system performance and functionality. The paper concludes to an advantage of using fuzzy logic as a formalism to predict the stock market. Reference [27] also addressed the challenge of price evaluation and decision making through a fuzzy-logic based system. They nominate a system that creates a fuzzy indicator, as they called it, that generates a buy hold or sell position. It is applied with the definition of certain fuzzy rules to express relationship among input indicators to the system. Similarly, Cheung and Kaymak [15] offered a decision model that incorporates the experience of trading experts, through deploying a fuzzy rule-based system. The authors assured success of the system with witnessing better risk-discounted returns. The use of fuzzy logic and its trading rules was continued by many other researchers [62], [16], [63]. Another important applied method along with fuzzy logic for optimizing and defining trading rules is genetic algorithms. Where, Allen and Karjalainen [5] established trading rules for the S&P 500 index. However, according to their paper, when compared to traditional buy-hold strategy, their strategy did not record any increase of return. However, in their paper [33] Fernandez-Rodregez *et al.* marked success with using genetic algorithms for optimizing trading rules.

The introduction of fuzzy logic deployment with financial market and securities evaluation is inevitably positive. Where, it replaced the usage of its rival in the domain the neural network. Neural networks had many limitations when deployed for handling technical analysis decision making. One of which is, its inability of handling uncertainty and ambiguity fairly contributing in such environment. Another, is their weakness with explaining the decision making steps followed, and the integration of trading rules. However, there exists in the financial markets the ability of using historical data prices of many well known and traded securities. Where, taking advantage of this added statistical knowledge in the learning phase of stock evaluation could give more promising results than using fuzzy logic exclusively. We adopt this point of view on what follows. Another potentially beneficial detail is the search of a different analysis technique than fuzzy inference systems and genetic algorithm, that could give importance to one side of information knowledge, and discarding another (a complete expression of trading rules could be very complex and probably impossible to reach).

3.3.3 Hybrid Artificial Intelligence Systems in Finance

The fact that all reasoning methods have their limitations drove the motivation of researchers towards the world of hybrid intelligent systems. In their paper [1], Abraham and Nath propose integrating different learning and adaptation techniques, to achieve synergistic effects through applying hybrid intelligence systems. They discussed the history of evolution of hybrid intelligence systems applied on different domains. They have proposed in another paper a hybrid intelligent system for stock market analysis. Where, they use neural network for a one day ahead stock price forecast and a neuro-fuzzy system for analyzing the predicted stock trend [2]. Hybrid models and approaches have been tackled widely in the financial market world of research. In the paper [47] Lin *et al.* developed a trading system model that predicts stock indices using a neuro-fuzzy framework. They applied the system on a stochastic oscillator type indicator. The applied system generated evident high returns in comparison with other investment strategies, such as neural networks and linear regression models. Many more neuro-fuzzy systems were adapted, for instance it was applied in [54] for financial time-series prediction, and in [14] for portfolios evaluation. In [6] and [40], also neuro-fuzzy systems have been used for greek and korean stock prediction.

Neural networks are mainly known for being accurate in prediction but, are considered very poor tool for handling uncertainty ambiguity and vagueness. Moreover, it completely lacks the ability of incorporating human emotions or handling the partial truth values between completely true and completely false. Therefore, Some other Hybrid systems were developed over the time for better responsiveness mainly, the probabilistic-fuzzy approaches. In their paper [67], Van den Berg *et al.* combined the interpretation of fuzzy logic with the statistical properties of probabilistic systems, through applying the Takagi-Sugeno (TK) probabilistic-fuzzy rule based model to analyze financial markets. They applied their proposed methodology to financial time series analysis. Assuming a given linguistic term set, they demonstrate how a probabilistic TS fuzzy system can be identified. An additional probabilistic-fuzzy system was put to use for estimating Value at Risk (VaR) that measures the expected loss of a portfolio [69]. Also, Teoh *et al.* [65] introduces a fuzzy time series model based on probabilistic approach and rough set rule induction for analyzing the stock market.

The integration of probability with fuzzy logic for analyzing the financial market and predicting price movements showed great contribution to the available models. Using the statistical powers of probability theory to deal with historical data available for securities of the financial market is definitely a winning added step. However most of the systems are fuzzy rule based systems. The accurate and sufficient modeling of such systems tend to be complex as mentioned in our earlier section 3.3.2. Also most of the studies do not include the effect of more than one individual technical indicator through there proposed model. By that, decreasing the beneficial effect of including the widest financial knowledge possible. On the other hand, using probability alone as a reasoning method to analyze the market, is fairly limiting. Since, probability theory is powerful with modeling the uncertainty regarding the market development and gives added value to the statistical learning of historical price data. However, there are other types of uncertainty present in the financial markets. Such as, the inevitable human effect on market efficiency and rationality, and the uncertainty in the definition of concepts and the human factor integrated with all aspects of the market. Human reasoning and actions is the more or less the basic factor of defining the financial market position and change from supply and demand till price change and market trending.

Therefore, this work motivation also follows the concept of a hybrid artificial intelligence system to take advantage of the most possible advantageous powers of available paradigms that deal with uncertainty in its various available types in this particular environment under study. In this manuscript, we propose a system that takes advantages of probability theory in dealing with statistical historical data, possibility theory competences in handling uncertainty and dealing with the available human factor, and the foreseeing capabilities of technical analysis with merging information from various technical indicators in the most efficient manner possible.

3.4 Possibility Theory

Possibility theory is mainly concerned with handling uncertainty, vagueness, and incomplete information. It was first recognized by Zadeh in 1978 [51] declaring that the theory can be used as a tool for uncertainty propagation with insufficient statistics or information knowledge.

3.4.1 Assumptions of Possibility Theory

The bedrocks of the theory were addressed and then extended by D.Dubois and H.Prade first in 1980 in their book [31]. Assuming S as a state of affairs, the mapping of S to a totally ordered scale L is the possibility distribution π, with top and bottom bounds of 1 and 0 respectively. The term π represents the state of knowledge of the agents about the state of affairs, in other words it represents the possibility or impossibility of the affair [30]:

$\pi(s) = 0$ serves that s is rejected as impossible;
$\pi(s) = 1$ serves that s is totally possible.

Possibility theory is known for specifying even minimal hypothesis without ruling out any state. The way of obtaining extreme forms of knowledge is the following:
Complete knowledge: for s_0, $\pi(s_0) = 1$ and $\pi(s) = 0$, $\forall s \neq s_0$ (only s_0 is possible)
Complete ignorance: $\pi(s) = 1$, $\forall s \in S$ (all states are possible).

Taking a basic example the query of the form "does event A occur?" where A is a subset of state, the necessity and possibility degrees are computed to obtain the answer to the query as follows (if possibility scale $L = [0,1]$):
Possibility degree: $\prod(A) = \underset{s \in A}{sup}\, \pi(s)$;
Necessity degree: $N(A) = \underset{s \notin A}{inf}\, 1 - \pi(s)$ Where $\prod(A)$ appraises to what extent A consists with π, while $N(A)$ appraises to what extent A is certainly implied by π. This duality is expressed by $N(A) = 1 - \prod(A^c)$, with A^c being the complement of A.

In possibility theory human knowledge is declared in an informative way, where it includes belief degrees attached to information. It is often known as the degree of certainty, often accompanied with constraints that should be abide. Assuming information A is certain to degree α, then the constraint becomes $N(A) \geq \alpha$. The possibility distribution representing this information is:

$$\pi_{(A,\alpha)}(s) = \begin{Bmatrix} 1, & if\ s \in A \\ 1 - \alpha & otherwise \end{Bmatrix} \tag{3.1}$$

This acts as a key to building possibility distributions. It is also possible to measure and include acquired pieces of information or evidence and updating $\pi_{(A,\alpha)}$. Another contribution of possibility theory is adding the typology of fuzzy rules, and making it possible to differentiate between rules that propagate uncertainty through reasoning, and rules that are just concerned with similarity-based interpolation. Also, conditioning and independence are tackled with possibility theory, where similar to Bayesian equations conditional possibility is defined: $\prod(B \cap A) = \prod(B|A) * \prod(A)$. Note that the possibility theory can be cast in either ordinal or numerical settings. In the numerical setting the * operation is considered as a usual product. While, in the ordinal setting the operation * is changed into a minimum. This is not the only form to define conditioning, in the numerical setting there are several other available ways to define it. It is also mandatory to mention an important example of possibility distributions, which is the fuzzy interval, where the calculus of fuzzy intervals is possibility-based. Possibility can be though of as halfway between fuzzy sets, probability, and non-monotonic reasoning [30].

3.4.2 Possibility Theory with information Fusion and Uncertainty Handling

The combination of possibility theory and aggregation operations of fuzzy set theory forms a helpful tool to deal with several source information fusion (such as the technical indicators presented in section 2.4, where the imprecision of information can be modeled by possibility distributions [29].

The fusion mode is usually dependent on the condition of the situation under study, there is not a mono-mode of merging that satisfies all situations, even within the same framework, like possibility or probability. There are various options to choose from, and it is important to keep in mind the difference between merging information, and aggregation of preference. The latter is more as a kind of filtering and estimation, as for fusion it is extracting the most reliable information out of imprecise data which is the case in this thesis proposed approaches.

Possibility-based information fusion is helpful with cases where pieces of information is poor in precision and uncertain, whether they are completely informative, or not informative at all. The technical indicators in our proposed analysis belongs fully to this framework of hypthesis. Thus, the purpose of the fusion is to find the most plausible value depending on the pieces of information available, in our case choosing the best decision to make depending on information coming from different technical indicators. The information to be used in fusion can differ in source, it can be coming from a sensor, a human, or a database. Thus, the information are heterogeneous and would also differ in type, they can be in the form of verbal linguistics, set of intervals, historical data series as in this thesis case situation, or any other form of information. Therefore, choosing the right fusion mode is much related to these mentioned criteria or conditions.

The information merging is a subject or reliability and truth, where the aim is to discard the wrong and keeping the right information hence, make the best use of available information. The available natural options of fusion are the logical combinations being the best candidates to actually do the job [29].

- **Conjunctive Combinations**, this fusion operation is mainly applied when all sources agree and are reliable, eg. the indicators lead to the same action.
- **Disjunctive Combinations**, this combination is used when sources disagree and at least one is wrong, or when unreliable sources are hidden under a reliable group of sources.

- **Quantified Fusion**, this technique is applied when all sources are known to be reliable, while it conbines their opinions conjunctively.

- **Prioritized Fusion**, this combination technique is obviously used when the sources are not of equal reliability.

- **Consistency-Driven Prioritized Fusion**, this fusion technique takes into consideration the consistency between sources. If the information between sources is consistent, then, the least reliable information is used as refinement. However, if the information of sources is conflicting, then the least reliable source information is discarded, eg, the indicators lead to different actions.

- **Averaging Operation**, this operation in information can be justified when the sources are considered as a single random source that produces different inputs.

These are the simple possibility fusion operations, that were introduced, along with other data fusion techniques, to cover some limitations in probability theory in dealing with some situations.

- Identifying one probability distribution might need information a lot more than what is actually available in the study, which causes problems of inefficient fusion.

- Information coming from experts are not only of limited reliability but also imprecise, therefore it is more real for the information to be presented as intervals rather that point values. Probability in general deals with random variables, but acquire limited ability with modeling imprecision.

- The consensus method is a voting-like procedure, which is a basic fusion technique. This method states that in case two sources give contradicting information, it proposes a mean of the two sources, without discarding the wrong one even in obvious cases. Therefore, as stated by possibility theory experts [29] the weighted average method feels more natural or realistic, where it offers a true answer instead of a preferred one.

- Another potential drawback in consensus method is assuming that all information stem from a single source. This is questionable is various situations, such as expert based information, or heterogeneous sensors.

- A limitation is also addressed corresponding to Bayesian method, which is the need of prior knowledge; where the analyst is sometimes considered an expert himself. However, this assumption is false in many cases where the analyst can only be conversant with reliability of experts, rather than the information itself.

It is exact that possibility adds some innovative solutions to some of probability theory limitation with pooling imprecise information. Nevertheless, it is believed that the type of pooling to choose for a certain problem solving or situation is a matter of context. There is no rule that is strictly followed to act as a universal pooling method that fits and applies to all situations. The choice of pooling depends not only on the described properties, but also on the degree of agreement or consistence between sources, and what is known about their reliability. For that purpose, in this thesis, proposed approaches will include various fusion techniques including probability and possibility based merging, to be judged according to their performance, and to be put into defiance under all possible conditions and changing parameters.

3.5 Conclusion

The purpose of this chapter is diving deeply through the research history on artificial intelligence and approximate reasoning in the world of finance, its contributions, strength and weaknesses. This aims to facilitate our perspective in spotting any drawback, limitation, or potential innovation in this field of study. Where, the reader's vision is directed towards the prospect propositions of this thesis work motivation. Then the following chapter delivers deeply and comprehensively the proposed fusion decision support systems, preceded by a pre-processing indicator based general system that processes data into a state preparing for the upcoming fusion decision support systems.

CHAPTER 4

Hybrid Probability Possibility Indicators-Based Decision Support Approach

4.1 Introduction

In the previous chapter, we have discussed in details the prior applied methods and techniques in dealing with financial markets along with the challenges faced and constraints integrated with the raised problematic. In this chapter, we introduce first the pre-processing general approach which is a hybrid probability-possibility preparatory system, that simply processes data into its fusion-ready state. Then, we propose multiple fusion approaches that are fed up with the output data of the pre-processing system. The proposed systems aim to overcome in performance individual indicators based analysis, and other applied fusion techniques. The systems deploy probability theory for its statistical claims, possibility theory for its history in dealing with uncertainty, and technical indicators being the voodoo of forecasting price change in financial markets.

4.2 The General Data Pre-processing System

Technical analysis as addressed earlier is mainly based on studying historical data to forecast the future of stock prices in the market, and the analysis process depends on multiple conditions and cannot be guaranteed with its success (refer to section

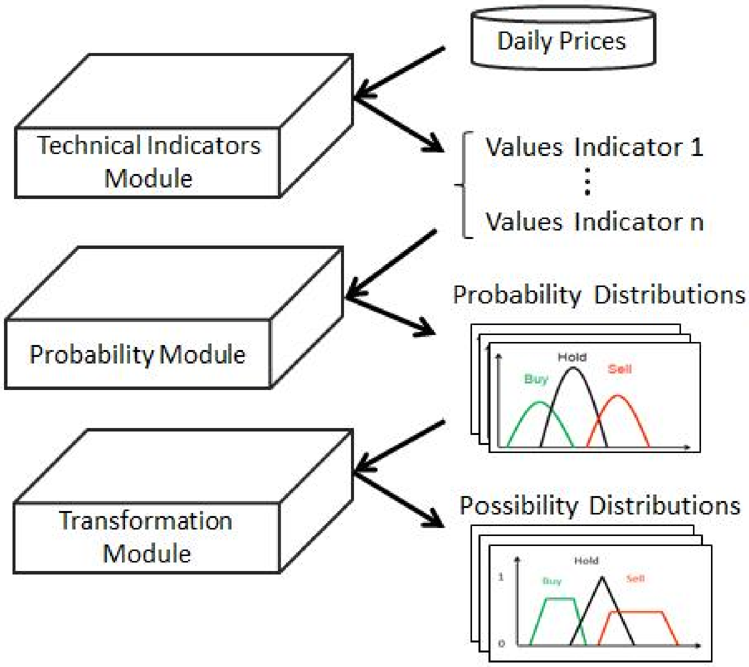


Figure 4.1: Flow Diagram of Probability-Possibility General Pre-processing System

2.3). As for the financial market itself, it has the limitation of being directly related to human behavior and emotions. This above studied conditions directed our work into deploying the powers of probability theory in dealing with historical data, where comes the defiance of the theory statistical claims in treating uncertainty. As for the incorporated human factor in the financial market, possibility theory claims to handle this kind of reasoning approximation. Therefore, various hybrid-probability possibility based decision support systems are proposed, evaluated and tested under various parameters to either validate or deny the superiority of the systems over typical individual analysis techniques. Prior to applying the decision fusion support systems a general approach was constructed to act as a pre-processing system for the data to be suitable for fusion. The system is basically divided into three modules:

1. Technical Indicators Module (TIM)
2. Probability Module (PrM)
3. Transformation Module (TrM)

As illustrated in figure 4.1, the system takes as an input time series with daily prices of any security to be examined, and it generates an output of possibility distribution functions representing the degree of membership to the decisions (Buy, Hold, and Sell) for the respective number of indicators used. This system is not a decision support

	A	B	C	D	E
1	Date	GOOG	CCI	EMA	SMA
2	6/23/2014	564.95	NaN	NaN	NaN
3	6/20/2014	556.36	NaN	NaN	NaN
4	6/19/2014	554.9	NaN	NaN	NaN
5	6/18/2014	553.37	NaN	NaN	NaN
6	6/17/2014	543.01	NaN	NaN	NaN
7	6/16/2014	544.28	NaN	NaN	NaN
8	6/13/2014	551.76	NaN	NaN	NaN
9	6/12/2014	551.35	NaN	NaN	NaN
10	6/11/2014	558.84	NaN	NaN	NaN
11	6/10/2014	560.55	NaN	555.7318	553.937
12	6/9/2014	562.12	NaN	556.8933	553.654
13	6/6/2014	556.33	NaN	556.7909	553.651
14	6/5/2014	553.9	NaN	556.2653	553.551
15	6/4/2014	544.66	-121.9	554.1552	552.68
16	6/3/2014	544.94	-100.42	552.4797	552.873
17	6/2/2014	553.93	20.3402	552.7434	553.838
18	5/30/2014	559.89	90.43836	554.0428	554.651
19	5/29/2014	560.08	80.45163	555.1405	555.524
20	5/28/2014	561.68	88.5764	556.3295	555.808

Figure 4.2: Technical Indicators Module Illustrative Example

system. It acts as the bedrock to the following proposed decision support systems in this dissertation. In what follows we describe deeply each module of the system individually.

4.2.1 Technical Indicators Module

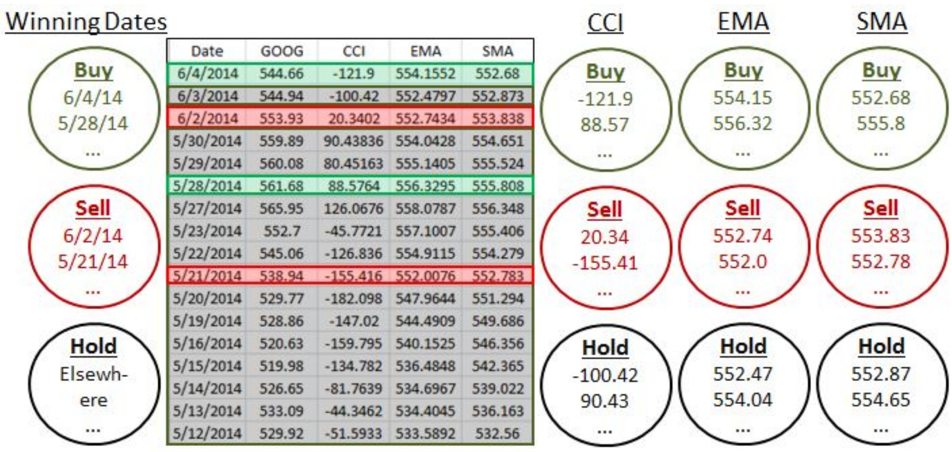
The first module of the system is the TIM. The role of this module is simple, it takes as input the historical data of any daily prices of a stock that are normally time series, and estimates following each mathematical formula of each indicator, the daily indicator values. For explanatory purposes, an example of introducing Google stock close price history of 20 days to the TIM, refer to figure 4.2. The values of three indicators CCI_{14}, EMA_{10}, and SMA_{10} according to formulas (2.7), (2.4), and (2.3) correspondingly. The values are estimated for the introduced daily historical data with selected periods indicated by the underscored numbers for each indicator. The appearance of NAN values (Not A Number) is due to the delay of the indicated period for the first calculated entry of each indicator. This way the TIM estimates the value of any number of indicators introduced to it according to its related mathematical expression.

4.2.2 Probability Module

Before feeding the data into the PrM module a step is applied for the purpose of recognizing from the historical data the winning dates of buying holding and selling past decisions. This is reached through examining price change, and as it is known if a trader buys at certain date d_t for a certain price and this price moves up in the future, then its a winning decision taken, and vice versa. If a trader sells before price falls, then it is a winning sell decision. Therefore the following logic was applied to distinguish past winning dates for the three decisions according to the following logic.

$$if\ p_{d+\gamma} - p_d \begin{cases} \geq \eta\% & then\ d\ Winning\ Buy \\ \leq -\eta\% & then\ d\ Winning\ Sell \\ Elsewhere & then\ d\ Winning\ Hold \end{cases} \tag{4.1}$$

Where p_d and $p_{d+\gamma}$ is price at date d and $d+\gamma$ respectively. Where, γ represents the selected number of days, and the $\eta\%$ is a percentage value chosen to represent non redundant price change. Applying this step returns winning past dates of buying, selling, and holding. introducing the previous module estimated Indicator values and the



Date	GOOG	CCI	EMA	SMA
6/4/2014	544.66	-121.9	554.1552	552.68
6/3/2014	544.94	-100.42	552.4797	552.873
6/2/2014	553.93	20.3402	552.7434	553.838
5/30/2014	559.89	90.43836	554.0428	554.651
5/29/2014	560.08	80.45163	555.1405	555.524
5/28/2014	561.68	88.5764	556.3295	555.808
5/27/2014	565.95	126.0676	558.0787	556.348
5/23/2014	552.7	-45.7721	557.1007	555.406
5/22/2014	545.06	-126.836	554.9115	554.279
5/21/2014	538.94	-155.416	552.0076	552.783
5/20/2014	529.77	-182.098	547.9644	551.294
5/19/2014	528.86	-147.02	544.4909	549.686
5/16/2014	520.63	-159.795	540.1525	546.356
5/15/2014	519.98	-134.782	536.4848	542.365
5/14/2014	526.65	-81.7639	534.6967	539.022
5/13/2014	533.09	-44.3462	534.4045	536.163
5/12/2014	529.92	-51.5933	533.5892	532.56

Figure 4.3: Illustration of Indicator Decision Winning Values Derivation

grouped winning dates of decisions, makes it easy to group for each indicator the winning decision values related to the derived dates. Figure 4.3 includes an explanatory scheme of how to group indicator values according to winning past decisions. Note that a testing of the above value parameters is discussed in the systems testing performance, section 4.4. Where, different combinations of (γ, η) values are tested for best performances. After grouping the winning values of each indicator for the three decisions, it becomes possible to make use of this already known past data to distinguish the indication of these values in the future. This step is mainly applied to take advantage of already knowing the past along with the statistical claims of probability with historical data. Therefore it becomes likely considered to estimate the probability density functions or distributions for the three decisions of each indicator, according to the estimated group of values earlier.

Density Estimate

The probability distribution of a continuous-valued random variable X is described with respect to its probability density function $f(x)$, where probabilities of X can be estimated according to the following formula,

$$f(x) = \lim_{\epsilon \to 0+} \frac{P(X < x + \epsilon) - P(X < x)}{\epsilon} \tag{4.2}$$

The main objective is normally estimating $f(x)$ from an observation sample of data $x_1, x_2, ..., x_n,$, in our case historical data of indicators. This parametric approach of estimating $f(x)$ is usually applied through assuming that it belongs to a certain family of distributions and accordingly calculating the data distribution. This approach is considered easy to apply, however it has a drawback of lacking flexibility. As for the non-parametric estimation, its idea is to avoid being restricted by a certain form. Some well known types of non-parametric density estimation types are histogram and Kernel Density Estimates (KDE). The histogram is a very convenient tool for measuring unknown probability density functions (pdf), but it lacks continuity. As for kernel density estimate, its is in many respects more preferable than the histogram, since it is capable of estimating a smooth continuous pdf and is also simple and easy to apply. In our work motivation, the indicators do not follow a standard practice or rule, yet they are mostly based on price and volume which are unknown random probability distributions. Hence, this eliminates the possibility of using any parametric estimation tool. As for the non-parametric density estimation in the technical indicators case, continuity plays an important role in measuring the relative entropy of indicators, which will be addressed in section 4.3.3 on the fusion approaches. Therefore, kernel density estimation is more likely to fit the needed probability distribution needed for

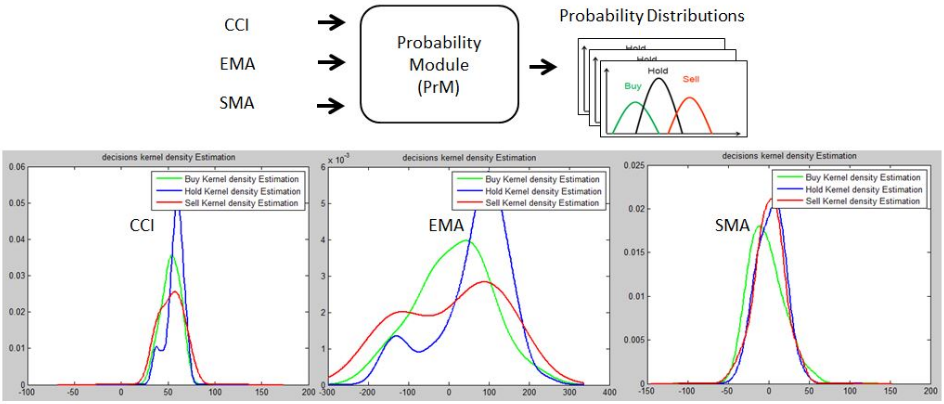


Figure 4.4: PrM Kernel Density Estimation

this work motivation, rather than histograms, and conditionally to the three possible actions (sell, hold, and buy) for CCI, EMA and SMA according to rule 4.1.

Kernel density estimation equation is represented as below, where the PrM uses the following estimation to generate for each indicator three decision pdfs according to the grouped winning values.

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{4.3}$$

Where K is the kernel density estimator of data sample $(x_i ... x_n)$, h is the bandwidth used for smoothing. The generally chosen to be equal to $N^{-0.2}\hat{\sigma}^2$, where $\hat{\sigma}^2$ is the variance estimate of the random variable. The Kernel Density function (KD) is a symmetric function that integrates to one, and n is the number of indicator values data samples forming the probability distribution, the more current kernel is the Gaussian one. Graphical illustration of indicator pdf estimation available through figure 4.4

4.2.3 Transformation Module

Because of the lack of a clear canonical way to directly construct the possibility membership functions, and since traditional probability has very good precision in the processing of historical data and uncertainty representation. The use of both theories domination techniques has become decisive. When Zadeh introduced possibility theory, it was considered as a counterpart to probability theory [51]. However, due to the lack of a precise method to generate possibility memberships for obscure information measurements, the method has not been much followed. For that purpose Dubois-Prade asserted that since a possibility measure can encode a family of probability measures, it is possible to develop transformation techniques to convert probability measures to possibility ones, and conversely. This has facilitated the use of both techniques in parallel, and encouraged the development of hybrid systems that take advantage of both theories simultaneously.

In their paper [28], Dubois and Prade described the most useful case of applying a probability to possibility transformation, as that of various week information sources availability or, a simpler computation with possibility than with probability. The two mentioned conditions are both satisfied in this domain conditions. First the imprecise information input comes from various indicators of different types. Second, the computation in a pure probability based system that uses many indicators can get very complex, as seen later in chapter 5.

The idea of inter-converting probability and possibility measures was addressed in the past but not by many scholars. Here is a more detailed comparative analysis on the subject [53]. In our study, the basic idea is to transform probability to possibility distributions. The notion of a relationship between probability and possibility distributions was first addressed by Zadeh on the theories consistency principle. Where, Zadeh stated that an event must be possible prior to being probable. Hence, a possibility degrees cannot be less than probability degrees in any case. The principal of probability-possibility consistency means that a probability measure P and a possibility measure $\prod$ are considered consistent if and only if $P \in P(\prod)$. As a reasonable refinement to the specificity ordering Dubois-Prade requests the satisfaction of the following constraint:
$\pi(x) > \pi(x')$ if and only if $p(x) > p(x')$. Where possibility distribution π is obtained from probability distribution p. Alternative principles have also been proposed, with the approach of Klir [41] being the most notable. It is based on the notion of information invariance with the following three assumptions:

- A scaling assumption forcing each value of π_i to be a function of p_i. Where π_i represents the possibility distribution at a certain value i.
- An uncertainty invariance assumption where, entropy $H(p)$ should be equal to information measure $E(\pi)$ contained in the transform π of p.
- Consistency condition of what is probable must be possible, should be satisfied by the transformation $\pi(u) \geq p(u)$, $\forall u$.

Klir's approach has limitations concerning the three assumptions. The uncertainty invariance assumption along with the scaling assumption together might reach a case where the consistency principle gets violated; refer to [31](pp. 258-259) for an example. Another limitation is the second assumption. Klir considers probabilistic and possi-

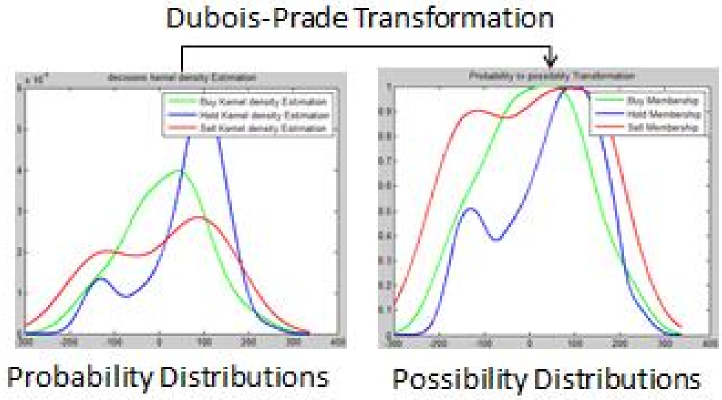


Figure 4.5: Distribution Transformation Example of CCI

bilistic measures consistent. Meaning that, entropy and imprecision collect the same type or facet of uncertainty.

For the above mentioned reasons, in this work motivation we apply a probability to possibility transformation following Dubois-Prade symmetric transformation techniques [28] where, we avoid such questionable pre-requisite assumptions. In the tranformation module, using the above introduced Dubois-Prade technique, we transform all indicators probability distributions into possibility distributions as illustrated in figure 4.5. The symmetric probability-possibility transformation $P \rightarrow \pi_i$ suggested by Dubois-Prade, was adopted in this module. It is defined by:

$$\pi_i = \sum_{j=i}^{n} min(p_i, p_j) \tag{4.4}$$

Where n is the number of indicator data samples used, and p_i and p_j are the probability estimates at indicator indices i, j. The purpose of this transformation is to deduce daily degrees of membership to the three decisions, buy, hold, sell, on a scale from 0 to 1 from each indicator, hence prepare the data for any later fusion processes.

4.3 Proposed Decision Fusion Support Systems (DSS)

The previous section introduced our proposed indicator-based pre-processing general system that is a system applied prior to any fusion to prepare data for the processing and fusion, note that figure 4.1 illustrates the pre-processing system modules, their input output, and functionality. As a fast briefing, the pre-processing system is fed up with historical prices of a certain security. It estimates indicator values from the

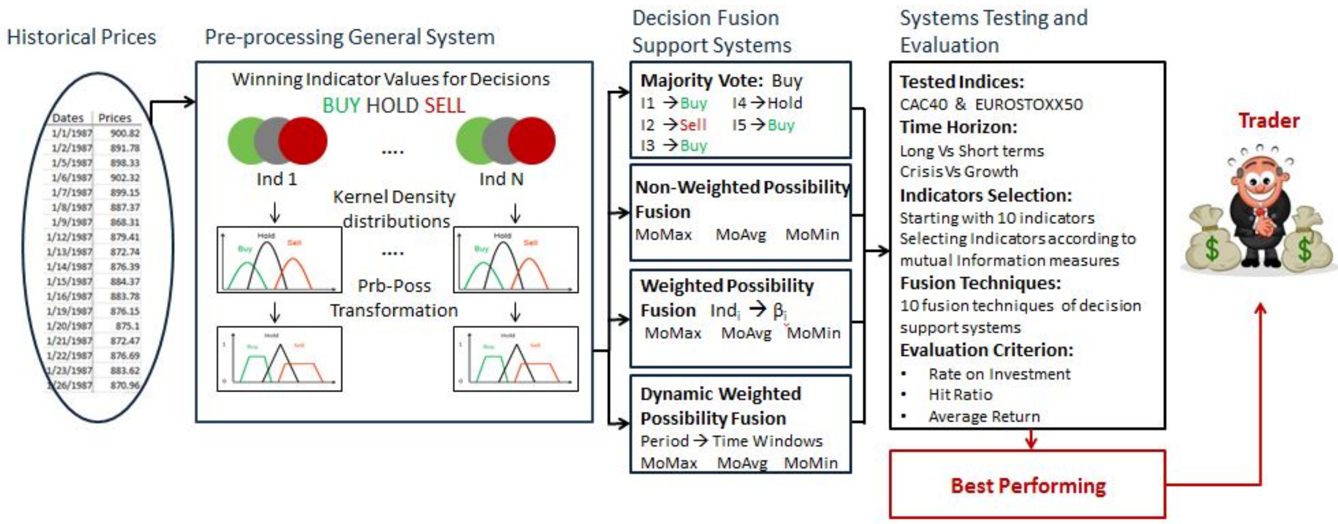


Figure 4.6: Briefing of Proposed Approach

historical price data and group indicator values into winning buy, hold, and sell values. Then, uses kernel density estimation to estimate the probability distributions of decisions for each indicator. In fact, the index distribution for each decision can be wither skewned or multi-modal (see figure 4.5). After that, the probability distributions are transformed to possibility distributions to be used later by the decision fusion support systems. A simplified scheme of the whole proposed approach with the decision support systems and the pre-processing system along with their positions and roles in the work-flow is shown in figure 4.6

4.3.1 Majority Vote Decision Support System

The majority vote decision system is the first proposed decision fusion system. It is mainly a very basic and instinctive type of fusion approaches. It is used in our study as a reference for comparison with the other more complex fusion systems. As its name indicates this system simply uses the majority vote of decisions recommended by N indicators used, or so explained as the most frequent decision of all indicators. However, it is important to know that this system is related to the domain of decision fusion, unlike the below presented systems that belong to the data fusion domain. Yet, the main purpose of all the proposed systems is to aid the trader in making a decision at a certain date. It uses the modules of the pre-processing general prob-poss system with an added module for majority vote, refer to figure for a fair understanding 4.7.
System work flow description:

- Generate from each indicator a degree of membership between 0 and 1, to the three decisions buying, holding and selling. This is achieved by mapping daily indicator values to the decision possibility distributions deduced by the earlier described pre-processing general system (following equations 4.1 4.3).

- Consider for each indicator the decision with the maximum degree of membership as the recommended decision by that indicator. Note that, in case two decisions have the same degree, the recommended decision is directly considered a hold.

- Choose the most frequent decision recommended by indicators as the majority vote, and thus the adopted decision at that date. Also note that,in case of equal frequency of decisions by indicators, holding is considered the adopted decision at that date.

Table 4.1 describes a simple day explanatory example, interpreting the work mechanism of the majority vote DSS, at a certain day d. Assuming that four indicators are introduced to the system (any number of indicators can be used), the earlier indicated steps are applied. In table 4.1 according to the maximum decision membership degrees, indicator 1 recommends selling for acquiring the highest degree of confidence. Similarly, indicators 2, 3 recommends buying and indicator 4 recommends holding. Therefore at date d buying is considered the adopted decision, for being the most frequent or majority vote of indicators.

4.3.2 Non-weighted Possibility Fusion Decision Support System

The non-weighted possibility fusion DSS also uses as basis the pre-processing general system modules, similar to the majority vote. However, this DSS has an important

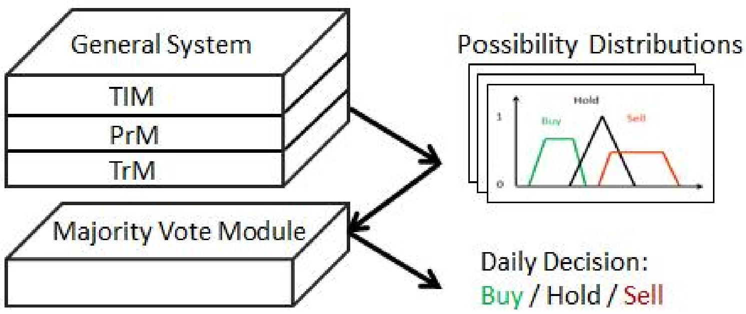


Figure 4.7: Schematic illustration of the Majority Vote DSS

Indicators	**Buy(d)**	**Hold(d)**	**Sell(d)**	**Recommded Decison(d)**
Indicator1	0.2	0.1	**0.7**	Sell
Indicator2	**0.6**	0.3	0.1	Buy
Indicator3	**0.8**	0.2	0	Buy
Indicator4	0.1	**0.5**	0.5	Hold
Adopted decision(d)				**Buy**

Table 4.1: Majority Vote Illustration Example

added role function, which is an actual type of fusion. Instead of primitively choosing the most frequent decision, it uses three possibility-based fusion techniques to generate a decision. The added module named the possibility module employs three possibility fusion techniques on the decision possibility distributions of multiple indicators, allowing the indicators to be better represented in the decision making process, illustrated in figure 4.8.
System work flow description:

- Generate decision membership degrees by mapping indicator values, as similarly explained in the majority vote DSS(following equations 4.1 4.3).
- Compute for each decision the Maximum, Average, and Minimum degrees of membership among that of the different indicators in use.
- Perform the following three fusion techniques: Maximum of Maximums (MoMaxs), Maximum of Averages (MoAvgs), and Maximum of Minimums (MoMins) through simply computing for each decision the maximum minimum and average of the degrees coming from different indicators and then finding the max degree of each

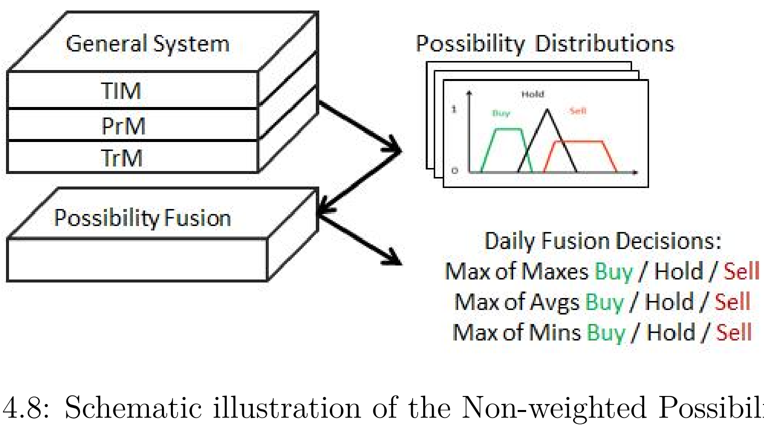


Figure 4.8: Schematic illustration of the Non-weighted Possibility DSS

Indicators	**Buy(d)**	**Hold(d)**	**Sell(d)**	**Fusion Techniques**
Indicator 1	0.2	0.1	0.7	
Indicator 2	0.6	0.3	0.1	
Indicator 3	0.2	0.8	0	
Maximum	0.6	**0.8**	0.7	MoMaxs(d)=Hold
Average	0.3	**0.4**	0.3	MoAvgs(d)=Hold
Minimum	**0.2**	0.1	0	MoMins(d)=Buy

Table 4.2: Non-Weighted Possibility Fusion Illustration Example

(maximum, minimum, average). The next paragraph details more the processing method for applying the fusion using the different techniques. Each technique will suggest a daily decision individually, which is simply obtained through computing the highest possibility degree of confidence (Maximum) among the above computed Maxes, Avgs, and Mins.

Table 4.2 describes a day sample interpreting the fusion techniques deployment of the non-weighted possibility fusion DSS, at a certain day d. Assuming that three indicators are introduced to the system, the decisions degrees of membership are estimated(Similar to the explained mapping technique in the Majority Vote Approach in section 4.3.1). The maximum, minimum, and average degrees of each decision are computed as shown in the table. Taking as example the Buy decision in the table it has three degrees from three indicators (0.2,0.6,0.2), the Max degree is 0.6, the Avg degree is 0.3, and the minimum degree is 0.2. After calculating the maximums minimums and averages, the highest degree of each is accordingly, selected to be the decision of each fusion technique (bolded values). At day d in table 4.2, MoMaxs and MoAvgs fusion techniques suggest holding, while MoMins suggests buying. Testing and evaluation of each fusion technique is addressed in the next section.

4.3.3 Information Theory: Entropy, Relative Entropy, and Mutual Information

In the section addressing power of possibility theory with information fusion and uncertainty 3.4.2, we have mentioned that not all sources of information are always equally reliable. Some information stem form the same source, hence, two source information might give similar information without adding actual new knowledge. For that purpose, it would not be subjective and scientifically logical to neglect the inclusion of this reliability matter to the fusion process. Before discussing the contribution of the next proposed approach it is important to have a look on mutual information and its relation with this work motivation and its integrated data nature.

Information Theory

Information Theory involves the quantification of information. It was introduced by Claude E. Shannon [20]. It has first been developed to model communication systems, and it has spread into reaching many areas including data analysis. The main use of such a theory in our research is related to determining the structure of dependencies among the set of variable (in our case technical indicators). There are different measures of information such as Entropy, Joint Entropy, Conditional Entropy, Relative Entropy (Kullback Leibler Divergence), and mutual information. The following measures are closely related, we take a brief look on the definition of each separately:

- **Entropy** It is simply a measure of uncertainty of a random variable X and a probability function defined by $p(x)$. Then, the entropy $H(x)$ of a discrete variable X is defined as [23]:
$$H(X) = -\int \mathrm{p(x)} \log \mathrm{p(x)} \, \mathrm{d}x \tag{4.5}$$
It is also possible to derive the definition of the entropy by certain related properties that a random variable must satisfy.

- **Joint Entropy** The prior defined measure is for a single random variable entropy, the following entropy is an expansion to that of two random variables X and Y. The definition of the joint entropy $H(X,Y)$ of a pair of random variables (X,Y) with a joint probability distribution $p(X,Y)$ is defined as follows [23]:
$$H(X,Y) = -\int\int \mathrm{p(x,y)} \log \mathrm{p(x,y)} \, \mathrm{d}x\mathrm{d}y \tag{4.6}$$
This implies that if X and Y are independent, then joint entropy is the sum of their individual entropies.

- **Conditional Entropy** The conditional entropy or uncertainty of random variable X given random variable Y, is simply the average conditional entropy over Y. Defined by the following equation [23]:
$$H(X|Y) = -\int\int \mathrm{p(x,y)} \log \frac{\mathrm{p(x,y)}}{\mathrm{p(y)}} \, \mathrm{d}x\mathrm{d}y \tag{4.7}$$
The definition of joint entropy and conditional entropy is provoked by the fact that the entropy of a pair of random variables is the entropy of one plus the conditional entropy of the other.

- **Relative Entropy (Kullback Leibler Divergence)** The entropy of a random variable can be described as the average amount of information needed to characterize a random variable. Here we introduce the relative entropy, which is a measure of the distance between two probability distributions. It can be introduced as an expected logarithm of the likelihood ratio. The measure of relative entropy $D(p||q)$ represents the inefficiency of assuming the distribution is q when the true distribution is p. It is defined by the following equation [23]:

$$D(p||q) = \int \mathrm{p(x)} \log \frac{\mathrm{p(x)}}{\mathrm{q(x)}} \, \mathrm{d}x \tag{4.8}$$

  The above measure of relative entropy is a non-symmetric divergence also known as the Kullback Leibler distance between the two probability distributions $p(x)$ and $q(x)$. Note that the relative entropy is non-negative and is zero in one case where, $p = q$.

- **Mutual Information** It is the measure of the amount of information one random variable holds about the other. It is the decrease in uncertainty of one random variable, according to the knowledge available about the other random variable. The mutual information is defined as the relative entropy between the joint distribution and the product distribution. For example, If we consider two random variables X and Y with joint probability density function $p(x, y)$ and marginal probability density functions $p(x)$ and $p(y)$, then the mutual information $I(X;Y)$ is defined as the follows [23]:

$$I(X;Y) = \int \int \mathrm{p(x,y)} \log \frac{\mathrm{p(x,y)}}{\mathrm{p(x)p(y)}} \, \mathrm{d}x\mathrm{d}y \tag{4.9}$$

  Mutual information can be considered a statistic for estimating independence between two random variables, and has a well-specified asymptotic distribution.

The reason behind introducing these different measures of information theory, is due to the need of its measures throughout the proposed approaches. As detailed in next section, relative entropy will be used for estimating a weight factor for each indicator, to be thus integrated with the fusion process for a more efficient and precise work mechanism.

4.3.4 Weighted Possibility Fusion Decision Support System

In the case of this thesis problem situation, in particular technical indicators, they are admitted to induce different knowledge about the financial markets. It is evident,

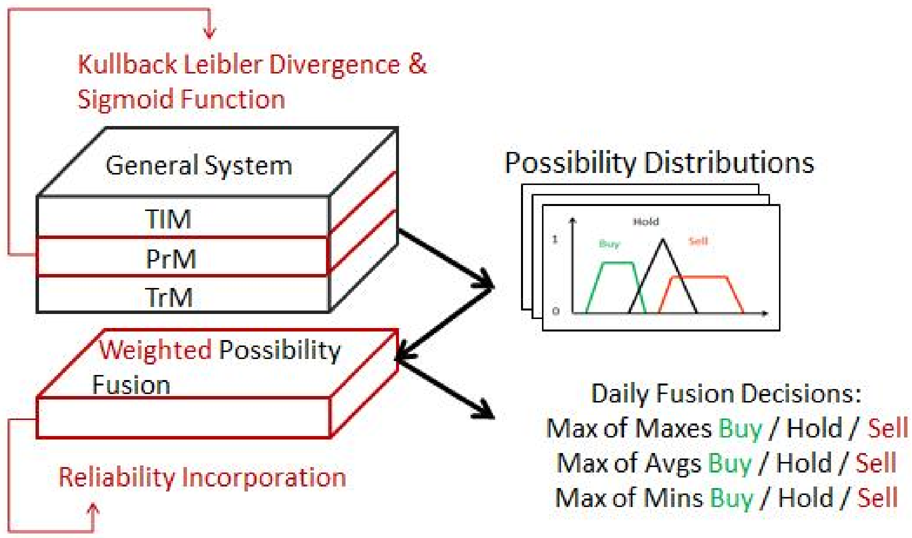


Figure 4.9: Schematic illustration of the Weighted Possibility DSS

as mentioned previously that indicators are of different types. Some indicators are based on volume therefore, they deliver data concerning the amount of people trading in the market. Other indicators imply knowledge about volatility, prices, and so on. Furthermore, identifying decisions out of indicators is a very important step, when analyzing prices. Very often when they are not well interpreted and used, indicators fail to imply winning decisions and might even lead to loss. For these reasons, it became evident to the proposed work mechanism the necessity to incorporate reliability to the fusion according the to the held information by each indicator. This urged the following proposed weighted possibility fusion DSS (see section 4.3.3).

With respect to the fusion process, this system similar to the previous presented DSS, also uses the three fusion techniques MoMaxes, MoAvgs, and MoMins. The novelty of this system lies in allocating for each indicator a weight factor, taking into account along the fusion process. This reliability incorporation during the fusion, gives more importance to the more robust indicators. The work mechanism of this system is similar to the earlier possibility fusion DSS, with two added steps. One step added to the probability module of the general system and another to the possibility fusion module, as illustrated in figure 4.9.
Additional Steps description:

- Kullback Leibler Divergence D_{KL}, also known as the relative entropy (as detailed earlier among the available information measures), is a measure of how different two probability distributions (over the same event space) are from each other. For the statisticians the probability distributions differ more or less according to the

difficulty in discriminating between them with the best test [43].
It is noted that the distance between the buy and sell probability distributions describes closely the robustness of an indicator. Since, the farther the distributions are, the easier their interpretation and the decision making process becomes. On the other hand, the closer the probability distributions of buying and selling the harder becomes the decision making and identification. This fact led our work direction into selecting the relative entropy measure as the best convenient information estimation measure to include this knowledge into our prior proposed fusion approaches. Thus, just after estimating the probability distributions in the PrM, a D_{KL} distance measure for each indicator is calculated. The KL divergence of probability distributions B (Buying), S (Selling) on a finite set x is defined as:

$$D_{KL}(B||S) = \int_{-\infty}^{\infty} \ln\left(\frac{b(x)}{s(x)}\right) b(x)dx \tag{4.10}$$

Where $b(x)$ and $s(x)$ denoting the densities of B and S.

- After the distance measure is calculated, the integration of the estimated measures with the possibility fusion process should take place. For this to be possible, it is important for the measure to meet the scale of the decisions possibility degree used in the fusion which is bounded between 0 and 1. Therefore, to change the scale of Kullback Leibler measures to meet the bound [0,1], Sigmoid Function is applied to the D_{KL} of each indicator. This allows the transformation of the earlier computed distance values in to reliability factors between 0 and 1, $0 < \beta < 1$.

$$\beta_i = \frac{1}{1 + e^{-\gamma(D_{KL})}} \tag{4.11}$$

Where D_{KL} refers to the Kullback Liebler Divergence measures, and γ could be any constant value as long as it is equally chosen for all indicator reliability factors.

Denoting π_i the possibility distributions of the indicators decision before incorporating reliability, and π_i' the possibility distributions of the indicators after including reliability, and is defined as $\pi_i' = max(\pi_i, 1 - \beta_i)$ Notice the two extreme conditions:

$When\ \beta_i \approx 0 \rightarrow \pi_i = 1\ \forall\ i, \quad \pi_i' = 1$

$When\ \beta_i = 1, \rightarrow \pi_i = \pi_i\ \forall\ i, \quad \pi_i' = \pi_i$
In other words, when the indicator is unreliable all the decisions are possible. While, for reliable indicators they are unchanged.

- The next step involves incorporating the reliability with the fusion application. This is achieved through adding a step prior to the typical used fusion techniques.

For each indicator, the maximum between its decision degrees and the level $1 - \beta_i$ is calculated. Then, the same fusion techniques are applied on the values of $max(Indicator_i, 1 - \beta_i)$. The purpose of this added step is basically discarding all non-efficient decision degrees of membership of an indicator by considering the level $1 - \beta_i$ as the new base for the indicators distribution, thus applying the fusion to the more efficient decisions.

Indicators	$1-\beta_i$	**Buy(d)**	**Hold(d)**	**Sell(d)**	**Fusion Techniques**
$MAX(Ind1, 1-\beta_1)$	0.3	0.3	0.3	0.7	
$MAX(Ind2, 1-\beta_2)$	0.2	0.6	0.3	0.2	
$MAX(Ind3, 1-\beta_3)$	0.4	0.4	0.8	0.4	
Maximum		0.6	**0.8**	0.7	MoMaxes=Hold
Average		0.4	**0.5**	0.4	MoAvges=Hold
Minimum		0.3	0.3	**0.4**	MoMins=Sell

Table 4.3: Weighted Possibility Fusion Illustration Example

Table 4.3 demonstrates the three fusion techniques with incorporating a reliability factor β_i for each indicator at day d. The degrees of membership are a result of choosing the maximum value between the level $1-\beta_i$ and the original decisions degrees of membership. The affected degrees in comparison with the earlier introduced example with non-weighted fusion in table 4.3 are marked in red. In the above example at date d MoAvgs and MoMins were the most effected by the added Indicators weight. The most affected is MoMins, with a complete change in its recommended decision from buying to selling. A detailed testing of performance, and evaluation, is tackled in the next section.

4.3.5 Dynamically Weighted Possibility Fusion DSS

In this DSS, an ongoing update of reliability is added, where the input is transformed into a sliding time window of a certain duration. Weighted possibility fusion is applied on each time window, the flow diagram is illustrated in figure 4.10.

The work flow description of this system, other than changing the systems fed input into a sliding time window of a predefined duration, is identical to that of the weighted possibility fusion DSS. The purpose of this inclusion, is giving more importance to recent data, and putting into defiance the effect of any change in the relative entropy of indicators with respect to time. Simply described, before the added factor of dynamism, the system did not take into consideration any effect of indicator reliability change with respect to any factor. And since, the reliability of indicators was noticed to be affected by time and not all the time static, it has become evident the necessity of including that fact to the proposed approach in this thesis, hoping that it would actually affect the overall revenue and efficiency of the system. A schematic representation of the changed reliability on possibility distributions is emphasized. The inclusion of reliability with the possibility distributions along three time windows is depicted in figure 4.11.

It is important to highlight that the following added step is considered as an innovative contribution to this approach, where dynamic reliability is a new concept added to the applied fusion approaches. Thus, this stands in side with the adoption of making a decision relying efficiently on various indicators instead of one. A detailed and complete testing, performance evaluation and analysis of all the above decision support systems is included in the next section.

4.4 System Performance Evaluation and Analysis

The above proposed decision support systems, are all tested under multiple varied criterion, putting into challenge all possible affecting parameters. This allows an accurate analysis, and a fair judgment of the suggested approaches, where transparency is abide and respected. Figure 4.12 includes a summary of all included criterion and parameters

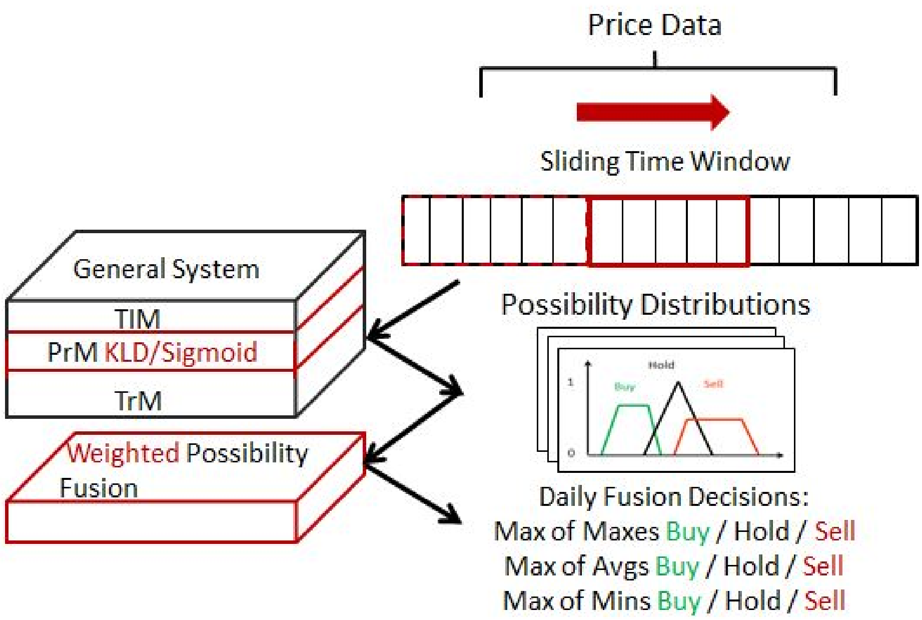


Figure 4.10: Flow Diagram of the Dynamically Weighted Possibility Fusion DSS

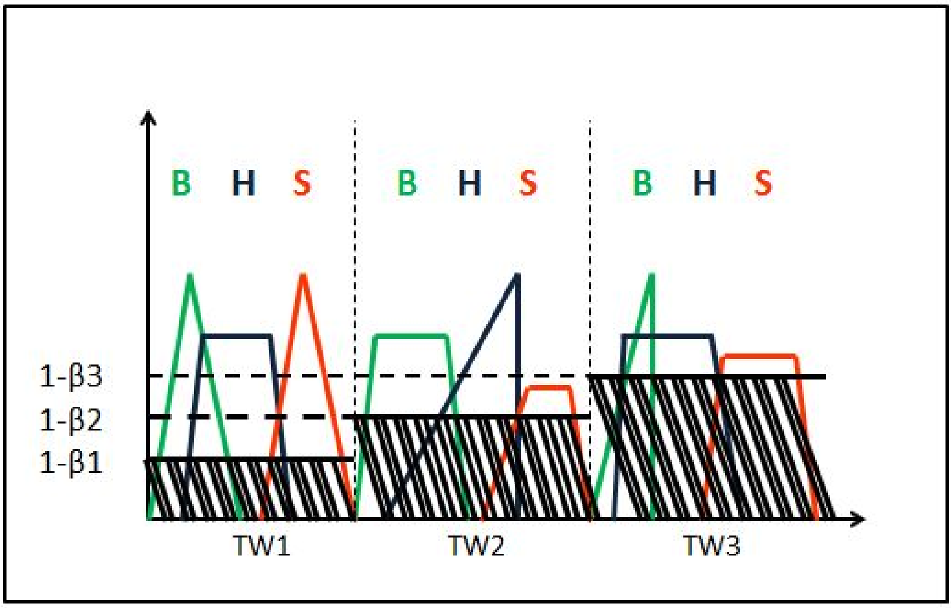


Figure 4.11: Schematic Representation of Dynamic Reliability on Possibility Distributions

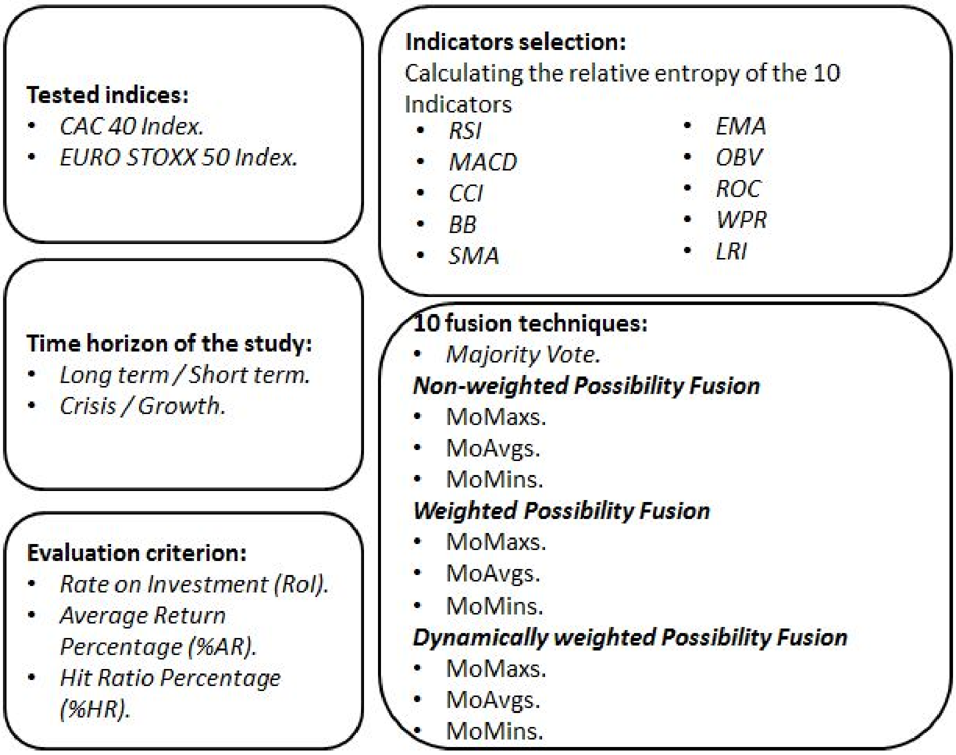


Figure 4.12: Testing Methodology Criterion Summarized.

affecting the system performances directly and indirectly, each to be explained later individually.

4.4.1 Tests on Indices

To test the performance of the different systems proposed, the two most common and traded European indices are used, the CAC 40 index and the EURO STOXX 50 index. Each deployed index is introduced below to enforce the reasons standing behind the selection of these indices:

- **Cotation Assistée en Continu 40, CAC40** is a very well-known stock index, the trading of which started in 1999. Since the Euro came to existence, the benchmark French stock market index CAC40, surpassed in terms of trading volume, the best performing options at the time the S&P500 and the DAX, becoming at one time the most traded index option around the globe. It represents the 40 most significant value stocks among the 100 highest market caps on the Paris Bourse [13].
- **EURO STOXX 50**, is a major barometer of financial markets in the Eurozone, therefore it is a indicator for the Eurozone market performance. It is the leading Blue-chip index for the Eurozone. EURO STOXX 50 represents the 50 most leading stocks of the 12 Eurozone countries [10].

4.4.2 Evaluation Criterion

In this work motivation evaluation process, a single index-trading is adopted as a start. The single index portfolio is a method to determine the right picking, in other words, making the right decision of buying, holding and selling. In fact, when evaluating a portfolio, it is difficult to distinguish the contribution of the picking (ie. choosing N shares among the market), and the allocation (ie. weighting the resource of the chosen shares). Therefore, in this approach the return or gain in price unit and in percentage is calculated upon each sell action preceded by a buy action throughout all testing periods. The performance evaluation of all systems is considered through calculating and comparing the Return on Investment RoI, the Average Return Percentage $\%AR$, and the Hit Ratio Percentage $\%HR$, of all systems.

RoI is used as a measure for the efficiency of an investment or to compare a number of investments. It is a very important factor to analyze for portfolio managers, since it represents the amount of gained money of an investment. It is an essential basic measure that is used for various financial purposes. It is used for investment evaluation, a companies financial behavior evaluation, managerial efficiency, and as a reference to establish the ceiling price in the regulated industries. It is calculated [64].

$\%HR$ is one of the key formulas for performance evaluation, it is a known important formula in mathematics. It is defined here as the ratio of the number of winning investments in the tested period. This measure is generally very interesting for decision makers, where they see it as a transparent means of judging performance of systems. It is calculated by dividing the number of the winning decisions that induced gain upon the tested period, over the total number of decisions made upon the tested period, multiplied by hundred.

The following values are calculated according to the following formulas:

$$RoI_{total} = \sum_{d=1}^{n} Selling\,Cost_d - Buying\,Cost_{d-\gamma} \tag{4.12}$$

$$\%AR = \frac{1}{n}\sum_{d=1}^{n} \frac{Selling\,Cost_d - Buying\,Cost_{d-\gamma}}{Buying\,Cost_{d-\gamma}} \times 100 \tag{4.13}$$

$$\%HR = \frac{N_{winning\,decisions}}{N_{total\,decisions}} \times 100 \tag{4.14}$$

Where, $Buying\,Cost$ is the price of the index at the buying date, $Selling\,Cost$ is the price of the index at the selling date, and n is the same as that of the equation of kernel density estimation and that of the dubois-prade transformation, according to equations (4.3) and (4.4) respectively. $N_{winning\,decisions}$ is the number of winning trades, and $N_{total\,decisions}$ the total number of trades made. It is important to mention, that these evaluation criterion are considered as a basic form of evaluation. In future work it would be appropriate to add different criterion that meet the different needs of different users, such as investors (buy-hold) traders(buy-hold-sell) and portfolio managers.

4.4.3 Studied Time Horizon

The testing is applied on daily data prices of each index during multiple varying time periods, short, long terms and growth, crisis times. The winning dates are estimated every 5 days ($\gamma = 5$), with detected 1% of change ($\eta\% = 1\%$). The different presented time horizons in table 4.4, allows us to identify the effect of different periods on certain parameters, such as the relative entropy, and efficiency of the market with respect to time. This permits a logical and transparent analysis of the systems behavior taking into consideration the time affected parameters of the proposed decision support systems.

Figure 4.13 illustrates the different performance of market indices upon times of crisis and growth, showing an opposite trending directions for both CAC40 and EURO

Studied Time Periods	**CAC 40**	**EURO STOXX 50**
Short Term	11 – 10 – 1999 till 11 – 03 – 2000	11 – 10 – 1999 till 11 – 03 – 2000
Long Term	02 – 01 – 1998 till 29 – 12 – 2000	02 – 01 – 1999 till 29 – 12 – 2001
Crisis	01 – 06 – 2000 till 01 – 06 – 2003	02 – 12 – 2007 till 02 – 12 – 2009
Growth	04 – 04 – 2003 till 04 – 04 – 2007	05 – 10 – 1996 till 05 – 10 – 1998

Table 4.4: Multiple Time Horizons Under Study

STOXX 50. This proves what has been discussed in the previous chapters about behavioral finance and the effect of humans behavior on the market (section 1.5.3), leaving it irrational, and not at all times efficient [66]. This fact as mentioned earlier encouraged our work motivation towards the usage of the hybrid prob-poss decision support systems, where the historical price data are handled during the pre-processing step with probability to help to derive the possibility distributions and allocate a relative reliability weight factor for each indicator. Also, taking advantage of possibility theory competences in handling the integrated human factor with market prices, for which, the information about the price is not completely known or considered.

Table 4.5, includes the results of estimating the reliability factor β_i of some indicators, to study the effect of time horizon on the relative entropy. It is clearly evident

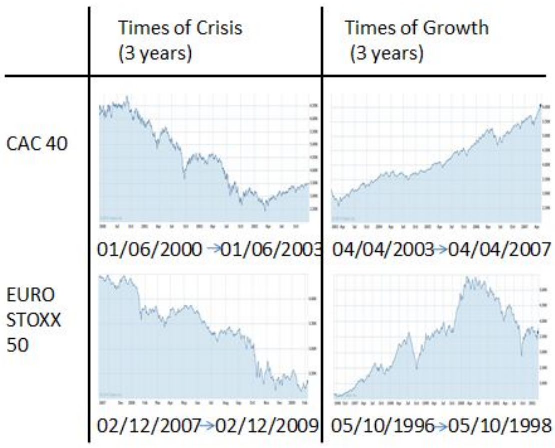


Figure 4.13: Inices Performance in Crisis Times Vs Growth Times.

Indicators	**CAC 40 Short Term (6 months)**	**CAC 40 Long Term (3 years)**
β_{LRI}	1	0.545
β_{RSI}	0.769	0.524
β_{MACD}	0.989	0.509
β_{ROC}	0.912	1

Table 4.5: Effect of Time Horizon on Entropy

that β_i of different indicators change with changed time horizons. Hence, the relative entropy of indicators is highly affected by time. Note that the estimation method is interpreted earlier in section of weighted possibility fusion decision support systems.

This effect of time on reliability is considered a very important step in the orientation of the work-flow. Due to this testing step and observation the whole idea of including a dynamic reliability was inspired. That is when the dynamic weighted fusion approach came into existence to model and integrate the effect of time horizon on both efficiency and relative entropy of indicators

4.4.4 Indicators Selection Process

The previous chapters addressed in details technical indicators, and mentioned the primary selected ten most popular and efficiency-known indicators along with their estimation and analysis techniques. For the testing and evaluation process, these ten indicators are exposed to a second selection process. Where, the indicators undergo a reliability test with both indices CAC40 and EURO STOXX 50 to be narrowed for simplification purposes, and for avoiding redundancy of information, into four of the most reliable indicators. In order to choose the best performing indicators, among the 10 introduced indicators, the relative entropy D_{KL} and accordingly the reliability factor β_i of each indicator is estimated. Identically similar in estimation to the reliability estimation technique introduced in the weighted possibility fusion DSS. As interpreted earlier, equations 4.10 and 4.11 are respectively applied on each indicator daily values, for both Indices CAC40, and Euro STOXX 50, permitting an accurate evaluation of indicator robustness. As noted in table 4.6, the four most robust indicators which gave higher reliability factors than the other indicators are: ROC, LRI, MACD, and BB. Note that the maximum reliability of each indicator among both indices (CAC40,

Indicators	EUEO STOXX 50	CAC 40	Maximum Reliability
β_{LRI}	0.578	0.545	0.578
β_{MACD}	0.526	0.524	0.526
β_{RSI}	0.515	0.509	0.515
β_{ROC}	0.513	1	1
β_{BB}	0.510	0.518	0.518
β_{EMA}	0.508	0.517	0.517
β_{OBV}	0.507	0.508	0.508
β_{CCI}	0.506	0.507	0.507
β_{SMA}	0.505	0.513	0.513
β_{WPR}	0.503	0.514	0.514

Table 4.6: Indicators Reliability Factors with Both Indices

EUROSTOXX 50) is computed and used for the selection process. The market in red indicators are selected to be used in the systems evaluation process.

4.4.5 Systems Performance Evaluation Results

For the evaluation process, the evaluation criterion defined earlier, $\%AR$, $\%HR$ are applied on individual indicators vs. different fusion techniques. The testing is applied on daily price data of both EURO STOXX 50 and CAC 40 on a historical time period of 3 years.

The results in table 4.7, represents the different decision support systems performance compared to the selected individual indicators performance. The test for each decision support system is applied following the earlier detailed estimation steps relatively.

The red highlighted values represent the fusion techniques that overcame the performance of the best performing indicator individually according the average return and hit ratio percentages. It is evident that MoAvrgs fusion technique of non-weighted, weighted, and dynamic weighted DSS, showed marked success over all indicators including the best performing. Nevertheless, there are also other techniques that also overcame the individually best performing indicator like Momins of the non-weighted fusion DSS. Also the MoMins and Momax along to the Moavgs of dynamic weighted DSS marked success in overcoming all individual indicators.

Individual Indicators vs. Fusion Techniques		EURO STOXX 50		CAC 40	
		%AR	*%HR*	*%AR*	*%HR*
ROC		0.3110	69.1	0.4511	65.1
LRI		0.6942	67.4	0.2791	58.1
MACD		1.2491	76.7	0.5861	61.9
BB		0.4941	41	0.7014	60
Majority Vote		1.067	77.2	0.6721	61.5
Non-Weighted Possibility Fusion	*MoMaxs*	0.3726	58.1	0.3002	58.3
	MoAvgs	1.5026	85.4	1.0023	68.7
	MoMins	0.8834	72.7	0.4491	59.1
Weighted Possibility Fusion	*MoMaxs*	0.3441	57.1	0.3002	58.3
	MoAvgs	2.0778	82.5	0.5704	66
	MoMins	0.5575	63.9	0.3551	59.1
Dynamic Weighted Poss-Fusion	*MoMaxs*	1.3251	68.5	0.8488	60.2
	MoAvgs	3.0034	85.7	1.4486	70
	MoMins	1.9240	67.5	1.2920	65

Table 4.7: Individual Indicators Vs Decision Support Systems Performance

It is noted in figure 4.14 that for CAC 40, *MACD* marked the highest cumulative gain over the studied period. While, for EUROSTOXX 50 *BB* marked the highest cumulative gain among studied indicators.

Since the MoAvgs fusion techniques of the three proposed possibility fusion approaches recorded higher performance than most other techniques with respect to $\%AR$ and $\%HR$, their cumulative gains are plotted in figure 4.15 along with the highest performing Indicator for each index, for comparison purposes. Figure 4.16

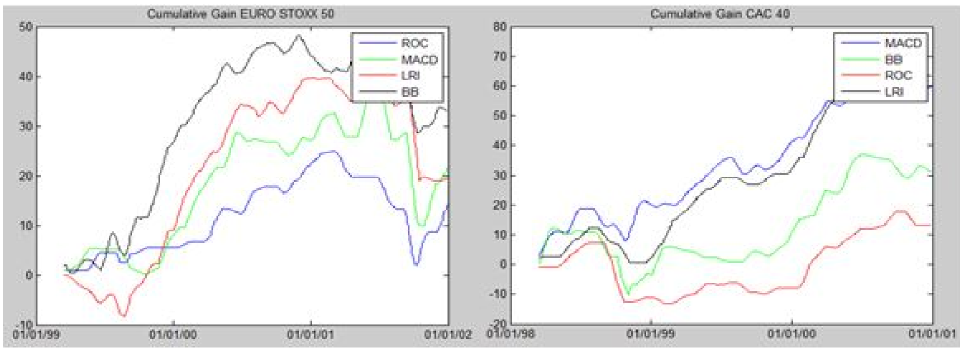


Figure 4.14: Comparing Performance of Indvidual Indicators Cumulative Gain Vs. One Another for both CAC40 and EUROSTOXX50

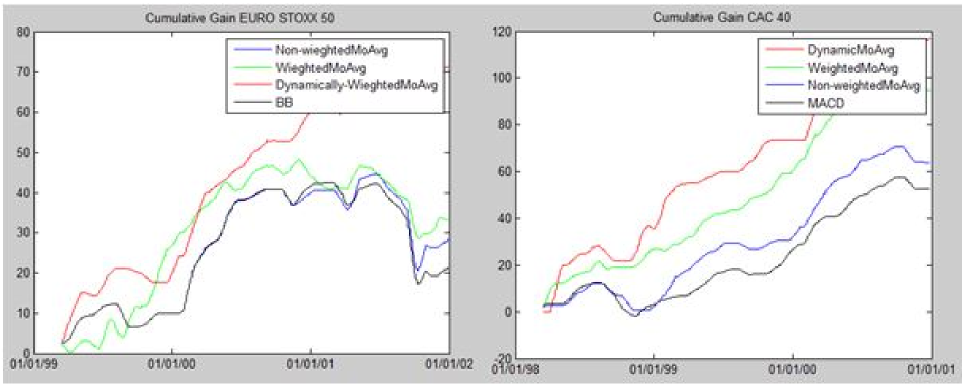


Figure 4.15: Comparing Cumulative Gain MoAvgs Fusion Technique Vs. Best Performing Indicators of Both Indices

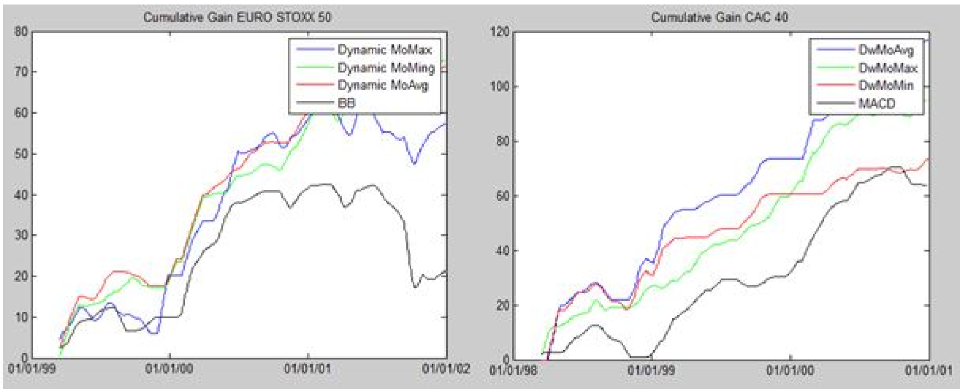


Figure 4.16: Comparing Cumulative Gain of Three Dynamic Fusion Teqhniques Vs. Best Performing Indicators of Both Indices

shows a plot comparing cumulative gain of the dynamic reliability fusion techniques MoMaxs, MoAvgs, and MoMins, to study whether they also overcome other techniques and best performing indicator with respect to cumulative gain as succeeded with $\%AR$, and $\%HR$, keeping in mind the dynamic weighted DSS possibility fusion techniques, are noted to be the most performing and efficient fusion techniques with respect to performance.

It is evident that multiple techniques proved to overcome the performance of the best performing indicator. The maximum of averages fusion technique of the three possibility fusion DSS recorded noted success. Surprisingly, even majority vote showed a high performance with respect to $\%HR$ on EURO STOXX 50, but not on CAC 40. The most robust fusion techniques according to multiple tests are the three dynamically weighted possibility fusion techniques, with MoAvgs being the best performing, with respect to all measures. Therefore we can declare that, all contributing DSS did meet expectations, where adding reliability and then dynamism to it did have a counted positive influence on the overall gain and hit ratio.

4.4.6 Winning Dates Testing

Another necessary parameter to be tested is the winning dates calculation parameters, related to the general pre-processing module addressed earlier in this chapter (refer to section 4.2.2). The probability module of the general pre-processing system includes the process of estimating the winning past decision dates of historical data. It is performed through analyzing price changes after a certain number of days, and checking whether the price increases or decreases by a specified price percentage. The two parameters of this estimation: number of days and percentage of price change are tested with different combination. Before analyzing the test results, we refer the reader to the previous introduced logic used for the winning dates estimation 4.1. Therefore, the performance test is applied on three different combinations of the (γ, η) measures.

The Test was applied on 6 years of daily historical prices of the Euro StoXX 50 Index (form January-1994 til December-2000). The three (γ, η) parameters combinations tested are as follows:

1. (30 days, 2%)
2. (60 days, 3%)
3. (90 days, 4%)

The Indicators used for the testing are those selected by the indicator selection process according to reliability tackled earlier in section (4.4.4)

1. Rate of Change (ROC)
2. Linear Regression Indicator (LRI)
3. Moving Average Convergence Divergence (MACD)
4. Bollinger Bands (BB)

The fusion techniques chosen for the testing are as follows:

1. Majority Vote (MoMax, MoAvg, MoMin)
2. Non-Weighted Possibility Fusion (MoMax, MoAvg, MoMin)

Individual Indicators vs. Fusion Techniques		(30 days,2%)		(60 days,3%)		(90 days,4%)	
		%AR	*%HR*	*%AR*	*%HR*	*%AR*	*%HR*
ROC		0.5519	60.6	0.3511	66.7	0.8761	70.6
LRI		0.6172	60.3	0.3112	54.2	0.3264	50.2
MACD		0.2067	60.1	0.8439	52.8	0.8977	63.6
BB		0.9809	59.0	0.9786	60.3	1.3518	60.2
Majority Vote		0.9189	69.1	1.0598	57.2	2.3674	57.7
Non-Weighted Possibility Fusion	*MoMaxs*	0.4000	60.0	0.3014	54.6	0.5270	58.0
	MoAvgs	1.0662	71.7	1.5116	70.6	1.7395	47.1
	MoMins	0.6356	65.6	1.5346	68.4	5.0168	71.4
Weighted Possibility Fusion	*MoMaxs*	0.4112	59.9	0.3014	54.6	0.5270	58.6
	MoAvgs	0.9869	69.7	1.2174	67.4	2.5352	47.1
	MoMins	0.6622	63.9	1.6535	59.1	10.248	80

Table 4.8: Studying the Effect of Different (γ, η) Combinations

3. Weighted Possibility Fusion (MoMax, MoAvg, MoMin)

As noted the Dynamic Weighted Fusion approach was excluded from this testing since its sliding time window of 80 days necessitates a minimum of 20 years of daily historical prices with expanding the duration of the sliding time window to be valid for testing with the three chosen γ measures (30, 60, 90). We can notice the difference in excellence between Average return and hit ratio. A high average return for a certain indicator or fusion technique does not necessitate a high hit ratio, and reciprocally. The reason behind this is because sometimes one winning trade inducing high gain, could effect greatly the return of the system, even when the system does not have much winning trades. Furthermore, a system might include many winning trades and few losing ones, but the winning trades do not induce very high return. Therefore, according to the need of the analysts, they can chose the most interesting measure to evaluate.

As mentioned when introducing the evaluation criterion (section 4.4.2), portfolio managers and analysts are more interested in the gain factor, therefore $AR\%$ is a more interesting measure in their field of study. As for decision makers $HR\%$ seems more interesting, since it better evaluates the system's performance in general regardless of the gained amount of money.

In order to make the analysis of table 4.8 easier we marked in red the measures of the best performing indicators with respect to Average Return and Hit Ratio. Moreover, we marked in blue the measures of fusion techniques that overcame that of the best performing indicator for each studied (γ, η) combination. The first noted event

Studied Techniques	**Portfolio Manager Best Solution**	**Decision Maker Best Solution**
Indicator	Bollinger Bands (BB)	Rate of Change (ROC)
Majority Vote	(90 days,4%)	(30 days,2%)
MoMax NPF	Not Indicated	Not Indicated
MoAvg NPF	(90 days,4%)	(30 days,2%)
MoMin NPF	(90 days,4%)	(90 days,4%)
MoMax WPF	Not Indicated	Not Indicated
MoAvg WPF	(90 days,4%)	(30 days,2%)
MoMin WPF	(90 days,4%)	(90 days,4%)

Table 4.9: Evaluation Results for Protfolio Managers and Deciosion Makers

in the testing is that Momax fusion techniques for both weighted and non-weighted possibility fusion did not show any superiority in performance over the best performing indicator. The second very interesting value is the extremely unprecedented high result of Weighted MoMins fusion technique when used with (90 days,4%) parameter combination, giving an $AR\%$ = 10.248% and $HR\%$ = 80%.

For a fair evaluation, we will address Average Return (interest of portfolio managers) and Hit Ratio (interest of decision makers) separately, in order to meet the needs of all analysts interests. The evaluation results are introduced in the form of a table for simplifying the outcome.

Table 4.9 sums up the evaluation outcome of table 4.8, where it shows the best solution or environment to use for portfolio managers and decision makers according to the results $AR\%$ and $HR\%$ respectively. First, we can notice the absence of results for both Non-weighted possibility fusion (NPF) MoMax and Weighted Possibility Fusion (WPF) MmMax for its fail of any superiority as noted earlier. For index portfolio managers the best performing indicator is Bollinger Bands (BB), since it overcame its competing indicator with respect to AR% for all three (γ, η) combinations. Similarly Rate of Change (ROC) is considered the best performing indicator for decision makers.

Studying the results for fusion techniques they all showed the highest AR% on the (90 days, 4%) parameter combination. Thus, for portfolio managers it is best to use the (90 days, 4%) parameter with all fusion techniques, while for decision makers the recommended parameter best to be used, is not constant. For both MoAvg fusion techniques with (30 days, 2%) recorded the highest HR%. And, both MoMin recordedthe highest HR% with (90 days, 40%), therefore considered the best to use.

4.5 Conclusion

The hybrid probability-possibility approach won indeed the bet of handling the vagueness, ambiguity, and uncertainty that accompanies the irrational markets of finance. Hence, combining both probability and possibility theories for achieving our purpose is a good proposition for strengthening the weakness points of each theory by complementing it with the other. Summing up the results, some decision support systems have shown better performance than expected. The complete scheme of testing should be enough to give all analysts an idea about which parameters and indicators to choose when applying the approach for guaranteed success. It is also very interesting that none of the fusion approaches induced loss, even though the less innovative techniques were defeated by other more efficient techniques like weighted and dynamic weighted fusion techniques. Another important thing to mark is the very inevitable additive effect of including dynamic reliability factors to the fusion (with MoAvg dynamic fusion technique being the most successful), and the significant accomplishment of weighted MoMin when used with (90 days, 4%) winning dates estimation parameters on both scales of Average return and Hit Ratio. Moreover, the current approach is to be compared with a pure probability system using Bayesian Networks in the next chapter to study the real effect of a hybrid approach versus a single theory dependent mechanism (ie. probability theory).

CHAPTER 5

Technical Indicators Learning for fusion with Bayesian Networks

5.1 Introduction

The preceding chapter introduced our contribution of a hybrid probability-possibility decision fusion approach. A complete transparent and fair testing of the proposed systems proved an innovative success over typical indicator analysis techniques used by analysts, investors, and traders with respect to profit, level of performance, and time consumption. The approach also included an analysis and learning that satisfies the needs of all concerned individuals from portfolio managers to decision makers. However, to strengthen our applied research it is important to diverse the testing strategies, to be as complete and convincing as possible. For that purpose, we introduce in this chapter a different purely probabilistic fusion approach with Bayesian Networks. The reason behind choosing Bayesian Networks is for its ability to model the influence of different random variables on each other, through studying the conditional dependence between them. The aim behind constructing the probabilistic directed acyclic graphical model (Bayesian Network), is to compute the probabilities of the decision (buy, hold, sell) knowing the probabilistic influence of indicators on it, via the Bayesian Network.

5.2 Graph Theory

Whenever an application with Bayesian network is addressed, the first notion integrated would be the probability directed acyclic graphical model. Generally, any application of Bayesian network dealing with real world challenges is directly related to a field lying between probability and graph theory. In this section we introduce the basics of graph theory, its terminology and structure properties, and its graph components.

5.2.1 Basic Terminologies: Graphs, Nodes, Arcs

Graph theory is the field of modeling objects through graphs. Mainly it is a structural model that illustrates relationships between different objects, for example technical indicators and buy, sell, hold decisions. The main basic components of a graph are nodes and arcs, where different nodes represent different objects under study and the arcs or lines between nodes represent the connections between these objects. There are two main types of graphs, directed and non-directed. As the names imply, a directed graph includes arrows (directed-arcs) from one node to another symbolizing the direction of causality or influence between the two related nodes. While, a non-directed graph simple includes lines as the non-directed arcs that connects nodes [19]. A mathematical representation of a graph can be expressed as $G = (V, A)$, where the component V represents the vertices or nodes of the graph, and component A is the finite set of pairs of vertices which is known as arcs, or edges. generally, an arc $a = (u, v)$ is defined as a pair of two neighboring nodes. When an arc is directed it contains an ordered pair of nodes (u, v) with a direction between them lets say directed from u to v represented as follows $(u \to v)$, where u here is the tail node and v is the head node. Taking the case where the arc is non-directed (non-ordered pair) its pair of nodes is represented as follows, $(u - v)$ with a line instead of an arrow. Often, each graph type has its mathematical representation. Where a directed graph has the above used representation $G = (V, A)$ (A if for Arc), an non-directed graph is denoted by $G = (V, E)$ (E is for Edge), and the mixed graph that includes both directed and non-directed edges is denoted by $G = (V, A, E)$ (has both Arcs and Edges). Figure 5.1

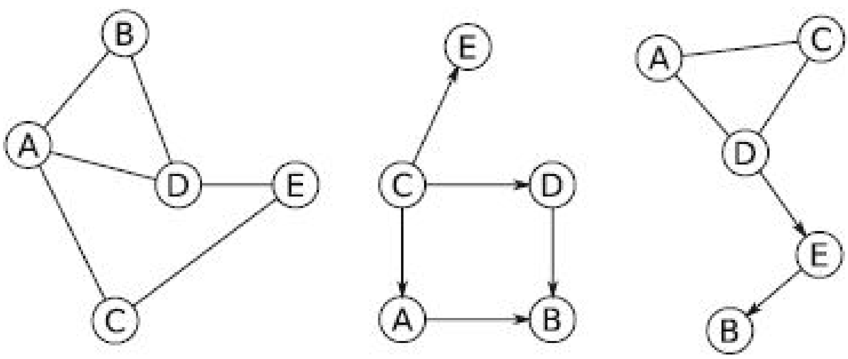


Figure 5.1: Non-directed, Directed, and Mixed Graph Structures

demonstrates graphs of the three graph types. Taking as example the three graphs in figure 5.1

1. **First graph (non-directed graph)**
   - The set of nodes in this graph is $V = \{A, B, C, D, E\}$, and the edge set is $E = \{(A - B), (A - C), (A - D), (B - D), (C - E), (D - E)\}$
   - The arcs are non-directed therefore notion $(A - B)$ and $(B - A)$ are exactly the same.
   - Also each connected nodes are considered adjacent nodes. Refer to [26] for more information of non-directed graphs.
2. **Second graph (directed graph)**
   - The nodes set is $V = \{A, B, C, D, E\}$ this graph has an Arc set not an edge set, and it is defined as $A = \{(A \rightarrow B), (C \rightarrow A), (D \rightarrow B), (C \rightarrow D), (C \rightarrow E)\}$.
   - Here the direction of arcs matter thus, $A \rightarrow B$ is different than $B \rightarrow A$. However, the graph is acyclic, therefore, only one arc between two nodes can be present. For further related explanation on directed and partially directed graphs refer to [8].
   - Also the two connected nodes are considered adjacent, and since its a case of directed graph one node is the head and the other is the tail according to the direction.
3. **Third graph (mixed or partially directed graph)**
   - This graph includes a combination of Edges and Arcs therefore, it is represented with two sets, $E = \{(A - C), (A - D), (C - D)\}$ and $A = \{(D \rightarrow E), (E \rightarrow B)\}$.

It is always possible to construct a non-directed graph from both directed and mixed graphs. It is simply performed by replacing the directed arcs by line edges, this type of graphs is often called a skeleton graph presenting the basic bedrocks of a graph. The interest in this chapter is exclusively with directed graphs since as we mentioned Bayesian Networks is in the form of an acyclic directed graph. In fact, our goal is to discover how indicators influence the decision, as well as the information shared by these indicators. Therefore, it is mandatory for our study to deal with directed graphs that can help us achieve our goal.

5.2.2 Structure of the Graph

The structure of the graph depends on the arcs pattern in it. In a Directed Acyclic Graph (DAG) the arcs between nodes are considered distinct, where at most one arc is available between two nodes. This constraint is related to the rule of acyclicity, where having two opposite direction arcs between two nodes causes a loop, which contradicts with the acyclicity constraint. There are two extreme forms of a graph the empty and saturated graphs. An empty graph is simply a graph with no arcs, and a saturated graph is the opposite, where every node in the graph has a relation with all the other nodes of the graph. In reality a graph of certain objects falls between these two extreme cases.

The interesting point in a graph structure, that forms an important statistical property is the path term. A path is a route formed by directed arcs from one node to another passing through a number of other nodes in the graph, as seen below its describes how the information "circulate" between the technical indicators as far as the final decision. Normally it is the incident of a sequence of vertices on the arcs between them. The arcs forming a path are considered unique, where the path passes through each arc only once. In the field of our interest, the directed graphs, it is obligatory for the arcs of a path to be in one direction, starting with the tail vertex of the first arc in the path, and ending in the head vertex of the last arc in the path.

Another important property of the graph structure in a directed graph is defining the order of nodes in the acyclic graph, which is derived from the direction of arcs. Following this assumption, the root nodes are the first nodes having no parents (no incoming arcs), the leaf nodes are the last nodes having no children (no outgoing arcs). Also, when having two nodes A and B belonging to the same path, with A preceding B but not directly and is the sequence of the ordered nodes. This leads to the believe of considering A as the ancestor of B, and B as the descendant of A. Furthermore if this same path is constructed from one single arc then it is concluded that A is the parent of B, with B being the child of A. Therefore the structural knowledge of a graph is considered very important for revealing major statistical properties that can help learning and analyzing the objects concerned with the problem under study.

5.3 Basics of Bayesian Networks

Bayesian Network (BN) is considered the most important computerized key tool for learning probabilities. They are graphical models that help reasoning under uncertainty, very convenient to the domain of this thesis problematic study. We have mentioned in the earlier section 5.2 of introducing graph theory, that BN is basically a DAG containing probabilities of objects taking into consideration the causal connection between them. The graph nodes in the network represent the variables and the arcs represent the causal connection between these variables.

5.3.1 Concepts

Being a graphical structure BNs enable us to reason and learn about an uncertain situation or domain including random variables, $X = \{X_1, ..X_i, ..X_n\}$. The random variables are represented in the BN as a DAG $G = (V, A)$. Each node defined as $vi \in V$ in the DAG corresponds to a random variable X_i. Obviously in our case the random variables are represented by either indicator values or decisions. The nodes and connected by links representing the dependencies between the variables $X_i \rightarrow X_j$. Since BN is of the DAG type it follows the constraint of acyclicity, thus disallows any loops or cycles in the connected paths between nodes. Each node has conditional probability distributions quantifying the strength of relationships between nodes.

Choosing Nodes and Values

Building a BN can include many steps of learning and knowledge engineering. A detailed explanation will be included in coming sections of this chapter. However in this section we introduce a simplifying example of problem solving with BN to familiarize the concept for the reader. The objective behind using BN to perform another decision fusion approach, is to construct a graph with the decision being the final node. Thus, enable us to distinguish the information of indicators that are contributing to the decision making process. In particular, detecting the conditional dependence relationship as discussed in section 5.3.4. Below is a descriptive text on the problem taken from [42]:

Example Problem: Lung cancer: A patient has been suffering from shortness of breath (called dyspnoea) and visits the doctor, worried that he has lung cancer. The doctor knows that other diseases, such as tuberculosis and bronchitis, are possible causes, as well as lung cancer. She also knows that other relevant information includes whether or not the patient is a smoker (increasing the chances of cancer and bronchitis) and what sort of air pollution he has been exposed to. A positive X-ray would indicate either TB or lung cancer.

The logic of learning with BN: The first step to start with is establishing the variables involved with the problem. In order to do that, it is important to know what nodes to include in the network, and what states can each node have. The values must be mutually exclusive and exhaustive for the discrete valued variables available. Meaning that, variables can take one value at a time. The possible types of discrete nodes are:

- *Boolean Nodes* this type of values can take two states True or false, 1 or 0. In the following network the Cancer node for example takes a boolean value where it can eaither be true (Cancer) or false (No Cancer).
- *Ordered Nodes* this type of nodes can take an order of states, for example the Pollution node in the following network can take three values or states $\{low, medium, high\}$.
- *Integral Nodes* this type of nodes can take values belonging to an interval, for example if we have an Age node representing the age of patients, then, it can take values from the interval $[1, 120]$.

The main important challenge, is to choose efficiently the type of states each node can have according to the needs of each situation. As a preliminary choice of nodes

Node Name	**Node Type**	**Node Values**
Pollution	Binary	*{low, high}*
Smoker	Boolean	*{T, F}*
Cancer	Boolean	*{T, F}*
Dyspnoea	Boolean	*{T, F}*
X-ray	Binary	*{pos, neg}*

Table 5.1: Lung Cancer Example: Choices of Nodes and Values

and their values, we begin with the restricted number of nodes available in that table 5.1. Choosing the nodes to represent the graph can be either efficient and complete or limiting and concise. For example, much of the diseases involved in the studied problem are not covered by the choice of nodes present in the table and according to that some limitations might appear in the network following its planning and choice of nodes and values.

Building the Network Structure

The second step for building a network would be choosing the right structure. If one node causes the other, then there should be a directed link representing this relationship

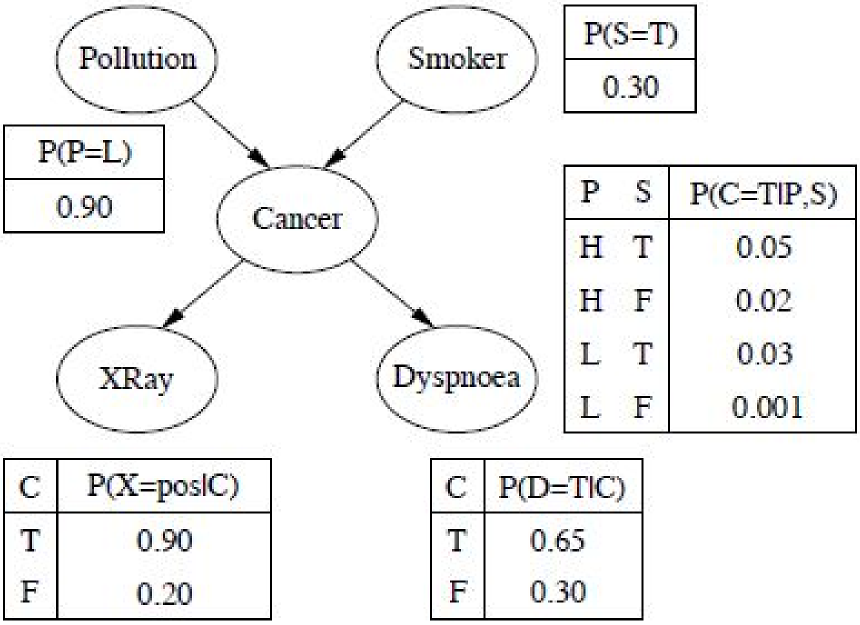


Figure 5.2: Bayesian Network Structure for the Lung Cancer Probelm

according to the direction of causality. For example, the value of an indicator can be derived from the other indicator value. Therefore starting to build the structure of the above given example, we should start with finding which factors can affect the patient's probability of having cancer. Similarly, we find the factors in the body that are affected by cancer. We then set the arcs and links. The output structure is shown in figure 5.2, where, pollution and smoking can cause cancer, and having cancer affects the patients ability to breathe and also affects the chances of having a positive result for X-rays. We can notice in the network that the cancer node has two parents, pollution and smoker, while X-ray and Dyspnoea are ancestors of smokers and pollution and children of cancer. Another important concept to derive from the structure is the Markov Blanket of a node. The Markov Blanket of the node consists of its parents, children, and other parents of its children. We have talked in the previous section 5.2 about the category of nodes, where the node with no parents is the root node, and the one with no child being the leaf node (for instance the decision node in our case) and all nodes in between are intermediate nodes. In the concept of causality, this can be interpreted as the causes and the effects. For example in the lung cancer problem we can conclude from the network that Pollution and Smoker are the cause of Cancer. While Xray and Dyspnoea are the effect of cancer. This facilitates the examination and interpretation of the network and it learning.

5.3.2 Joint Probability Distribution

The Bayesian network is mainly deployed for capturing the problem being modeled through following what is assumed to be an efficient structure that models all interfering factors of the problem. For more understanding we consider a BN with n nodes X_1 to X_n, taken in that order. The value in the joint distribution is represented in its simplified form as $P(x_1, x_2, ..., x_n)$. There is a way to factorize joint probabilities through following Chain rule of probability theory, Applying this rule we obtain [42]:

$$\begin{aligned} P(x_1, x_2, ..., x_n) &= P(x_1) \times P(x_2 \mid x_1)..., \times P(x_n \mid x_1, ..., x_{n-1}) \\ &= \prod_i P(x_i \mid x_1, ..., x_{i-1}) \end{aligned} \tag{5.1}$$

As already mentioned in section 5.3.4, the value of a certain node in the BN structure is conditional on the values of its parent nodes only. Therefore, the above equation reduces to:

$$P(x_1, x_2, ..., x_n) = \prod_i P(x_i \mid Parents(X_i)) \tag{5.2}$$

Where $Parents(X_i) \subseteq \{X_1, ..., X_{i-1}\}$. Taking as example the lung cancer problem, the joint probability expressions of the system can be simplified as follows:

$$\begin{aligned} &P(X = pos \cap D = T \cap C = T \cap P = low \cap S = F) \\ &\qquad = P(X = pos \mid D = T, C = T, P = low, S = F) \\ &\qquad\quad \times P(D = T, C = T, P = low, S = F) \\ &\qquad\quad \times P(C = T \mid P = low, S = F) P(P = low \mid S = F) P(S = F) \\ &\qquad = P(X = pos \mid C = T) P(D = T \mid C = T) P(C = T \mid P = low, S = F) \\ &\qquad\quad \times P(P = low) P(S = F) \end{aligned} \tag{5.3}$$

In order to simplify this chain of conditional probabilities, we have to define the notion of conditional independence.

5.3.3 Conditional Independence

In order to understand the working mechanism of BNs, it is important to understand the connection between the network structure and its conditional independence.Figure 5.3 illustrates three figures belonging to different node causal structures in a BN, we tackle each one of them separately.

Causal Chains

Figure 5.3 (a) shows a causal chain of three nodes A, B, and C. Where A causes B which causes C. This causal chain induces conditional independence as follows:

$$P(C \mid A \cap B) = P(C \mid B) \tag{5.4}$$

The above equation states that the probability of C knowing A and B is the same as probability of C knowing only B. This is because if we already know B has occurred, then knowing A does not give any information about C. If we want to verify this application on the lung cancer problem in figure 5.2, we can say that the probability that a patient has Dyspnoea is directly related only to the probability of the patient having cancer or not without any needed extra knowledge about smoking or pollution state and probabilities. However, not having knowledge about the patient's cancer probability while knowing that the patient smokes increases both our beliefs that she has cancer and suffers from breathing problems (Xray). An example of chain rule is Dyspnoea being conditionally independent of the smoking knowledge knowing the probability of the patient having cancer.

Common Causes

In this v-structure, we have a common cause B for two variables A, and C, refer to figure 5.3 (b). Understanding the concept through the lung cancer problem, we can take as example cancer being the cause of two symptoms Dyspnoea and short breathing (positive X-ray). This generates a conditional independence structure similar to that of the causal chain as follows:

$$P(C \mid A \cap B) = P(C \mid B) \equiv A \perp\!\!\!\perp C \mid B \tag{5.5}$$

If there is no information about the patient's cancer state, then having knowledge about any of the symptoms would increase belief in having cancer, which therefore, increases the belief of suffering form the other symptom. While, if we know the probability of having cancer then knowing the probability of one of the symptoms wont have any significance on knowing the chances of other. Translating the case to our technical indicators into BN approach, a steady trend, eg. bullish or bearish, will similarly affect indicator values and consequently facilitate the decision making.

Common Effects

Here, as the name also indicates, we have an effect node with two causing nodes, just like the structure in figure 5.3 (c). In our case of technical indicators this is translated by the situation where we determine whether two indicators share a common effect being the decision. This structure of common effects rises a conditional independence structure opposite to that of common causes and causal chains. Where, in this structure the parents are dependent only when knowing information about the common effect. They are marginally independent without this information. We can say the parents are conditionally dependent, since they become dependent in the condition of having knowledge about their common effect.

$$P(A \mid C \cap B) \neq P(A \mid B) \equiv A \not\perp\!\!\!\perp C \mid B \tag{5.6}$$

To translate that on the lung cancer problem for more understanding, we take the node Cancer being a common effect for Smoking and Pollution. If information on one of the parent nodes (causes) is not available, then, this increases the belief in the other. Therefore not having any knowledge about whether the patient is a smoker or not (knowing he has cancer) increases his probability to being exposed to high levels of pollution.

After examining the relation between the structure and the conditional independence, we can easily imagine the effect of violating the order of causality on the amount of arcs available and thus on the complexity of the probability learning process.

5.3.4 Markov's Property and Conditional Probability

In order to be able to model a BN, it is first mandatory to assume the Markov Property. This property asserts that each variable is conditionally independent of its non-descendants given its parent variables. Note that it contains all the information of its parents. In our indicators problem, an indicator can gather all the information. BNs that follow this property are considered Independence maps or (I-Maps). Bellow is a listing of map type definitions.

- **Independence Maps (I-Maps):** If there is one-to-one correspondence between the nodes V and variables X in a graph G, then G is an I-map of the probabilistic

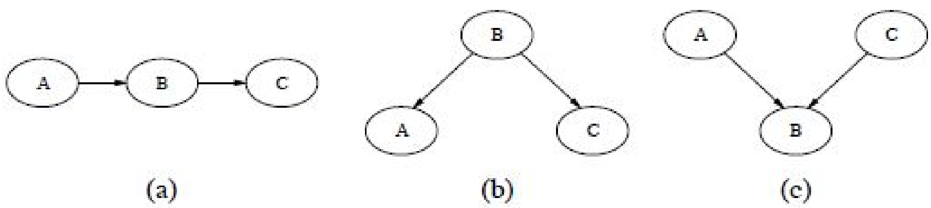


Figure 5.3: Causal Chains (a), Common Causes (b), Common Effects (c)

dependence structure P of X.

$$A \perp\!\!\!\perp_P B \mid C \Leftarrow A \perp\!\!\!\perp_G B \mid C \tag{5.7}$$

- **Dependence Maps (D-Maps):** Similarly, G is a dependency map of P if every arc in the graph belongs to a direct dependence in the system.

$$A \perp\!\!\!\perp_P B \mid C \Rightarrow A \perp\!\!\!\perp_G B \mid C \tag{5.8}$$

- **Perfect Maps (P-Maps):** G is considered a perfect map when it is both I-Map and D-Map.

$$A \perp\!\!\!\perp_P B \mid C \Leftrightarrow A \perp\!\!\!\perp_G B \mid C \tag{5.9}$$

Note that in the above equations $\perp\!\!\!\perp_G$, represent the graphical separation persuaded by the lack of a certain arc. Also, $\perp\!\!\!\perp_p$ denotes the probabilistic conditional independence in the system.

After completing the nodes, values and structure construction, comes the need to quantify the relationships between the related nodes. In the case of discrete values (like in the lung cancer problem) it is represented as Conditional Probability Table (CPT). To construct the CPT for each node, we examine all the possible combinations of the parent nodes (parent set instantiation). For each distinct instantiation we define the probability that the child will take each of its values. Taking for example the CPT of the Cancer node in the lung cancer problem, the parents of Cancer are Pollution and Smoker. Pollution can take on the values $\{Low, High\}$ and Smoker can be $\{True, False\}$. Thus, the CPT of Cancer includes all possible joint values of its parents, giving the following probability cases of having cancer <0.05,0.02,0.03,0.001>. The probability of not having cancer for the above cases is one minus the above probabilities <0.95,0.98,0.97,0.999>. The CPT for root nodes includes the prior probabilities, for example the prior probability of a patient being a smoker is 0.30. This comes from prior knowledge of statistics indicating the 30% of the patients are smokers, and 90% of the population is exposed to low levels of pollution. It is evident that the number of values in the CPT can get very large according to the number of parents of the concerned node, where it is exponential in the number of parents. For a boolean network the number of probability cases in a CPT of a node having n parents is 2^{n+1}. As for non-boolean networks it is K^{n+1} where n is also the number of parent nodes, and K is the number of states the node values can take.

5.3.5 D-separation

We have talked in the previous section about the connection between the conditional independence and the structure of node placement and its effect on belief change. We have seen examples on conditional independence when having knowledge about a certain node can activate or block a relationship between other nodes. For a portfolio manager this can be translated to the fact that an indicator value can block or activate another indicator value. In other words, either the information of an indicator is already brought by another one, or the information is complementary to the information brought by the other indicator. Another example related to the structure in figure 5.3 (a), where knowing information about node B blocks information of C relevant to A.

However this concept applies to sets of nodes as well as pairs. For more explanation, assuming the Markov property is satisfied, it is possible to find the dependency between two set of nodes say X and Y, given information about set of evidence nodes E. For achieving that goal we follow the criterion of direct-dependent separation, or what is known as D-separation. The concept behind D-separation is to connect dependence with connection and independence with separation.

- **Path** the definition of a path between two set of nodes X and Y is any sequence of nodes between any member node belonging to set X with any member node belonging to set Y. However, the following conditions should be satisfied, where a node is not allowed to appear in the sequence more than once, and every pair of adjacent nodes are connected by an arc.
- **Blocked Path** For explaining the blocked path we refer to figure 5.4. A blocked path occurs when for example we have a set of nodes E and a node Z belonging to a path that satisfies one of the three conditions below:
  1. if node Z belongs to set E and has a chain structure (5.3.3).
  2. if node Z also belongs to set E and has a common cause structure (5.3.3).
  3. if none of Z and its descendants belong to set E, while Z is indirectly a common effect of paths coming from E (5.3.3).
- **D-Separation** Assuming we have three set of nodes X, Y, and E. Set of node E is considered D-separating sets X and Y, if knowing E blocks all paths from any of nodes belonging to X to any of nodes belonging to Y. Therefore, given Markov property, we can say that if X and Y are D-separated by E, then knowing E makes them conditionally independent. Translating this case to the lung cancer problem for a better understanding, we assume that an observation of the Cancer node is

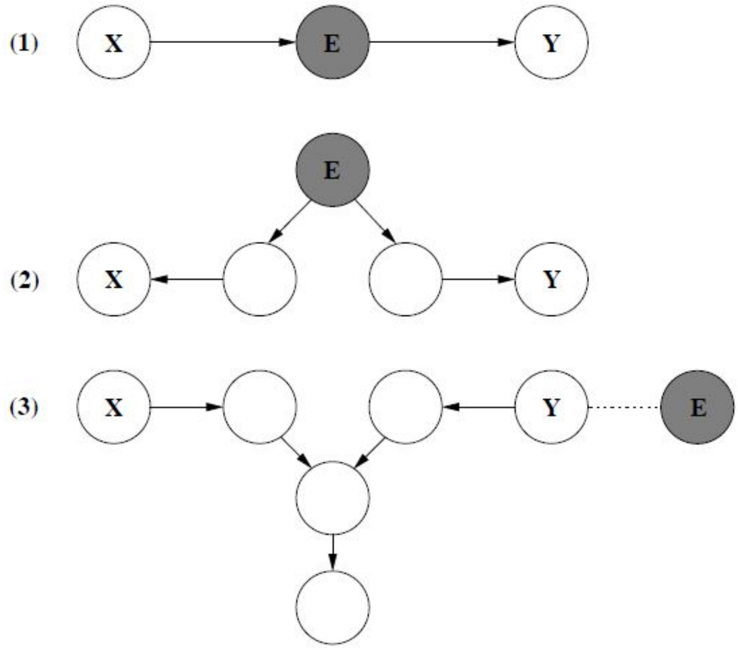


Figure 5.4: Three Conditions of a Blocked Path

our evidence. We can then say that P is D-separated from X and D according to first blocking condition. Also X is D-separated from D following the second blocking condition. However, S would have been d-separated from P where C is not observed (Condition three of blocking).

- **D-Connection** Two sets X and Y are considered d-connected when there is a path between their nodes not blocked given E

5.4 Reasoning with Bayesian Networks

After addressing the uncertainty and domain representation in a BN, it is time to inspect about the process of domain reasoning. The process of conditioning also known by inference, which is performed through a flow of information within the network

5.4.1 Inference

Bayesian Inference in a Network is a way for using Bayes' rule to update the probability estimate for a hypothesis when extra information or evidence is acquired. It can be described as a way to update the Bayesian Network with arrival of new incoming data concerning its variables, it is a very known and deployed technique in mathematical statistics. This process becomes particularly important in analyzing sequence of data in a dynamic manner, where it has been used and applied through a wide range of fields [18].

We start by proposing an explanatory basic algorithm for Bayesian inference. Given a set of competing hypotheses which explain a data set, then, for each hypothesis:

1. Convert or transform the likelihood and prior data information into probabilities.
2. Multiply them together.
3. Then, normalize the outcome in order to reach the posterior probabilities of each hypothesis given the evidence.
4. Choose the hypothesis that is most probable.

Bayesian inference estimates the posterior probability as a consequence of two antecedents, a prior probability and a likelihood function derived from a probability model for the data to be observed. Following is the formula for calculating the posterior probability following Bayes' rule.

$$P(H \mid E) = \frac{P(E \mid H) \times P(H)}{P(E)} \tag{5.10}$$

Where, H denotes the hypothesis, E is the evidence or new data that have not been considered in the prior probability estimation.

5.4.2 Structure Learning

The most basic way of constructing a BN is having the model specified by an expert and then simply applying inference. However, it is not normally so simple in real life, where, not always an expert knowledge is available, but mostly the network is learned from available data as the later proposed case of learning the network from the historical data of technical indicators. For that purpose, there has been proposed ways for learning the structure and parameters of a Bayesian network from the data available. There are two categories of structure learning algorithms for Bayesian networks, the constraint and the structure based algorithms.

- **Constraint-based algorithms** This is an algorithm that analyzes the probabilistic relations of the networks satisfying Markov property with tests of conditional independence (refer to 5.5.2 for detailed explanation). Where then, the graph that satisfies the D-separation statements is chosen and constructed. This algorithm is mainly based on the seminal work of Pearl on maps and its application to causal graphical models, in particular the Inductive Causation (IC). The IC algorithm is based on conditional independence tests for learning the structure of a Bayesian Network [55]. The steps of the IC algorithm are as follows:

1. finding pairs of variables that are connected by an arc regardless of its direction.
2. identifying the v-structures among pairs that are not adjacent, and with a common neighbor.
3. Identifying compelled arcs and choosing the direction for each recursively.

- **Score-based algorithms** While the score-based algorithm basically chooses the Network structure of that recording the highest score. Where in this learning algorithm each candidate network is allocated with a score. The score is a measure that reflects the superiority of the network with respect to its fit. A following step in this algorithm is maximizing the score of the network with using from among the various available heuristic search algorithm. Examples of such available algorithms are Greedy, Genetic and Simulated Annealing Search algorithms. [57] includes a detailed description of the available heuristics, with approaches from the artificial intelligence field.

5.4.3 Parameter Learning

After learning the structure of the Bayesian Network from available data, it becomes important to estimate and update the parameters of the distribution, which is simplified by the Markov property. In reality local distributions include few variables with bounded and non-scaled dimensions. This increases the problem of dimensions, which lead to the development of two approaches for parameter estimation. The first approach is based on the maximum likelihood estimation, and the second approach based on Bayesian estimation. However, since the conditional independence relationships encoded in the network structure fix large parts of the parameter space, the number of needed parameters for identifying the global distribution is reduced [60].

5.5 Learning Bayesian Networks with the bnlearn Package in R

In this section, we translate the power of learning with BN the studied problem of this manuscript. We start by introducing available model selection and parameter estimation techniques in classical statistical models. Then, we apply various learning techniques on our studied problem of technical indicators using the bnlearn package in R.

5.5.1 What is bnlearn and the Purpose Behind Using it

bnlearn is a package in the R statistical software tool. It provides multiple implementation techniques for learning the structure of a Bayesian network, along with optimization algorithms, score functions, and conditional independent tests. It has the needed versatility factor for handling a data analysis problematic. It is a relatively simple and easy to use tool, and that is because it provides a single object class for all its algorithms. It also includes a set of functions that allow the user to apply basic inference procedures and descriptive statistics.

There are other available packages developed for learning with Bayesian Network. One of them is the "pcalg" which is a package that applies the PC algorithm and specializes in causal interpretation in the network [39]. Another package is "deal" which implements the hill-climbing search algorithm for mixed data [25]. There are also the "gRbase", "gRain", and "gRc" packages that are responsible for implementing multiple inference methods.

5.5.2 Available Algorithms

We have mentioned in section 5.4 that there are two categories of structure learning algorithms: the constraint and score based algorithms. We have mentioned that all constraint-based algorithms are based on the inductive causation algorithm developed by Verma and Pearl [55] in 1991. Its main role is to provide a theoretical framework for learning the causal structure model. as for the score-based algorithms, they are simply parts of the heuristic search algorithms with having a score that defines the probability distribution of a network. The bnlearn package implements various algorithms of both categories.

- **Constraint Based Algorithms**

1. *Practical causation* (PC) this algorithm is the first practical version of the inductive causation algorithm application, it follows a selection process that moves in a backward manner. It starts from a saturated graph and goes forward through the selection.
2. *Grow-shrink* (GS) this algorithm is based on grow-shrink Markov blanket detection algorithm, which is normally deployed in structure learning.
3. *Incremental association* (IAMB) this algorithm is based on a two step selection plan, it follows the incremental association Markov blanket
4. *Fast incremental association* (Fast IAMB) this algorithm uses as its basis the IAMB algorithm introduced above. However, it is a variant of IAMB, which mainly reduces the number of conditional independence tests.
5. *Interleaved incremental association* (inter.iamb) this algorithm is also based on the IAMB and is a variant of it, its main additive role is avoiding false positives in the detection phase.
6. *Max-min parents and children* (MMPC) this another algorithm used in structure learning, it learns an underlying network with non-directed arcs. It is a forward selection technique that detects neighbors through maximizing the minimum association measure.

- **Score Based Algorithms**

  1. *Greedy Search* This a structure score-based learning algorithm that starts with an empty graph and explores the search space of the network. It aims to find the best structure, through adding and removing arcs while studying the score. When the score stops to increase the search is stopped and the network is selected. Algorithms following this time of search are random restarts, tabu, and Hill-climbing (available in the bnlearn package as hc)
  2. *Genetic algorithms* This algorithm is based on iterations that are repeated until the fittest network is established. This search process depends on crossovers and mutation imitating natural evolution.
  3. *Simulated annealing* This structure learning algorithm also depends on its search space on the score, through performing a stochastic local search with accepting and rejecting changes according to scores increase or decrease.

We offered above a brief introduction to all deployed structure learning algorithms and their availability with the bnlearn package in R. Now, in order to deliver better understanding to the reader, it is mandatory to examine a detailed description of some of the algorithms followed steps on learning. Bellow are two algorithms, the Inductive causation which is a highly used constraint-based structure learning algorithm, and the Hill-climbing one of the best functional score-based structure learning algorithms.

- **Inductive Causation Algorithm**

  1. For each pair of variables A and B in V, the algorithm looks for a set $S_{AB} \subset V$, such that A and B are conditionally dependent given S_{AB} and $A, B \notin S_{AB}$. If no set is found to follow these conditions, it places a non-directed arc between nodes A and B.
  2. For each non adjacent variables pair A and B having a common neighbor C, verify whether $C \in S_{AB}$. If the condition is not true change the initial non-directed path of nodes into the following $A \rightarrow C$ and $C \leftarrow B$
  3. Set direction to remaining non-directed arcs, through recursively applying the two rules below:
     (a) if A is adjacent to B, and there a non-directed path available between A and B, then set the direction between them to $A \rightarrow B$
     (b) if A is not adjacent to B and $A \rightarrow C$ with having a non-directed path between C and B, then set the direction between then into $C \rightarrow B$.
  4. Return the reached DAG whether fully or partially directed.

- **Hill-climbing Algorithm**

  1. Start with selecting a network structure G with set of variables V. It usually starts with an empty graph but, it is not necessary.
  2. Estimate the score of the graph G with the following $Score_G = \text{Score}(G)$.
  3. set $maxscore = Score_G$.
  4. keep repeating the two steps below as long as the $maxscore$ keeps increasing.
     (a) for every possible arc addition, deletion or reversal that does not contradict with the low of acyclisity :
        i. estimate the score of the changed network G^*,$Score_{G^*} = \text{Score}(G^*)$.
        ii. if $Score_{G^*} > Score_G$ set $G = G^*$ and $Score_G = Score_{G^*}$.
     (b) update $maxscore$ to the value of $Score_G$.
  5. Return the DAG.

5.6 Technical Indicators fusion Approach Learned with Bayesian Networks

In some cases of analyzing a Bayesian network, it is applied using pre-specified networks. This is usually followed when expert knowledge is available and the connection between variables and its direction is already delivered by domain experts along with posterior probabilities. However in real life, most of the situations do not have pre-specified networks. Furthermore, most of the time the knowledge of experts is not available. For that purpose in particular, the various structure and parameter learning techniques were developed in the first place. In our technical indicators problem, we have daily historical data (and thus estimated technical indicator values), which we can use for learning structure of our network.

The network will include the same ten indicators we started with for the hybrid probability possibility decision support systems approach introduced in details in section 2.4.2. The network will include also a node representing the decisions being the leaf node, therefore the overall pre-specified number of nodes in the network will be 11 nodes. We already have available daily price and indicator values. Also according to the previously proposed winning dates estimation approach we have a daily proposed winning decision of either holding selling or buying refer to section 4.2.2. The (γ, η) parameter are chosen of the same previous values of the long-term testing applied on the EUROSTOXX 50 $(5, 1\%)$. The reason behind this choice of parameters is to deliver a fair testing with these expressed in chapter 4. Our next step would then be using this data to learn the most optimal network to use for our testing.

5.6.1 Structure learning with bnlearn

We have introduced above the available constraint and score based structure learning algorithms. We attempt to learn our network with all possible algorithms available in bnlearn. The work environment used will be an integration of Matlab and R software.

Basic Approaches

We start by importing the estimated indicator values and winning decisions from matlab into the R environment, where the bnlearn package could be used. Withing the *bnlearn* package we have all needed structure learning functions that in their most basic form takes as argument the data frame including all variables in the model. Therefore, before applying different structure learning techniques, we bind the imported data into one variable using the *cbind* function in R to prepare data into the different learning algorithms below [59].

1. import the following data from Matlab: LRI, MACD, ROC, BB, WPR, CCI, RSI, EMA, SMA, Prices, Decisions.
2. Bind data together using the *cbind* function as follows:
   $>$ *data=cbind(LRI,MACD,ROC,BB,WPR,CCI,RSI,EMA,SMA,Prices,Decision)*
3. Apply the following constraint and score based structure learning functions using default argument values:
   (a) Hill Climbing Greedy Search Score Based Approach
   $>$ *bn.hc=hc(data)*
   (b) Tabu Greedy Search Score Based Approach
   $>$ *bn.tabu=tabu(data)*
   (c) Grow Shrink Constraint Based Approach
   $>$ *bn.gs=gs(data)*
   (d) Incremental Association Constraint Based Approach
   $>$ *bn.iamb=iamb(data)*
4. plotting the graphs and showing the learning records of each algorithm.

The output score and constraint based networks are illustrated in the networks below. We start by the score-based learning algorithms, in figure 5.5 of the output network we noice that the Decision node (colored in red) is directly affected by the price node only. While in figure 5.6 illustrating the tabu search, we can see that the decision is affecting the price and no other node is affecting the decision. Both graphs are fully directed where we can see in the records displayed in figure 5.7 that the number of non-directed arcs for both algorithms is zero and all arcs of both networks are directed. We can also notice that the tabu search took 722 tests for its learning process, while Hill-climbing took 405 tests for its complete learning.

For the constraint algorithms we can notice that the two networks of Grow-shrink and IAMB algorithms are identical, with a non-directed arc between prices and decision nodes. Even other tested constraint-based algorithms gave the same network structures as that of grow-shrink and IAMB illustrated in figures 5.8 and 5.9 respectively. Even changing the conditional independence test (available in the bnlearn package) would not make any difference and will give the same results of the two teste algorithms. As for the records of learning available in figure 5.10 we can see that both algorithms

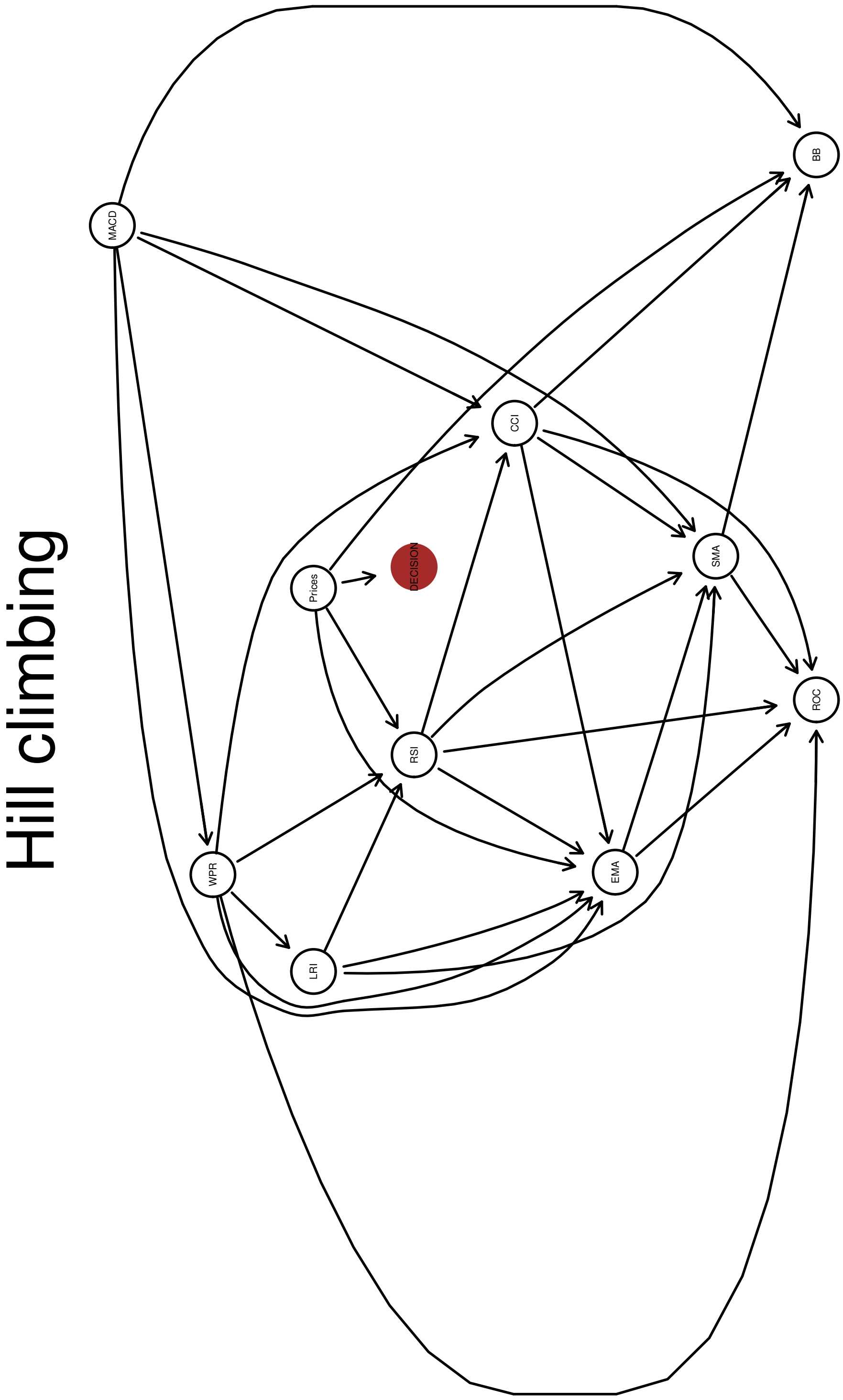


Figure 5.5: The Hill-climbing Greedy Search BN Structure

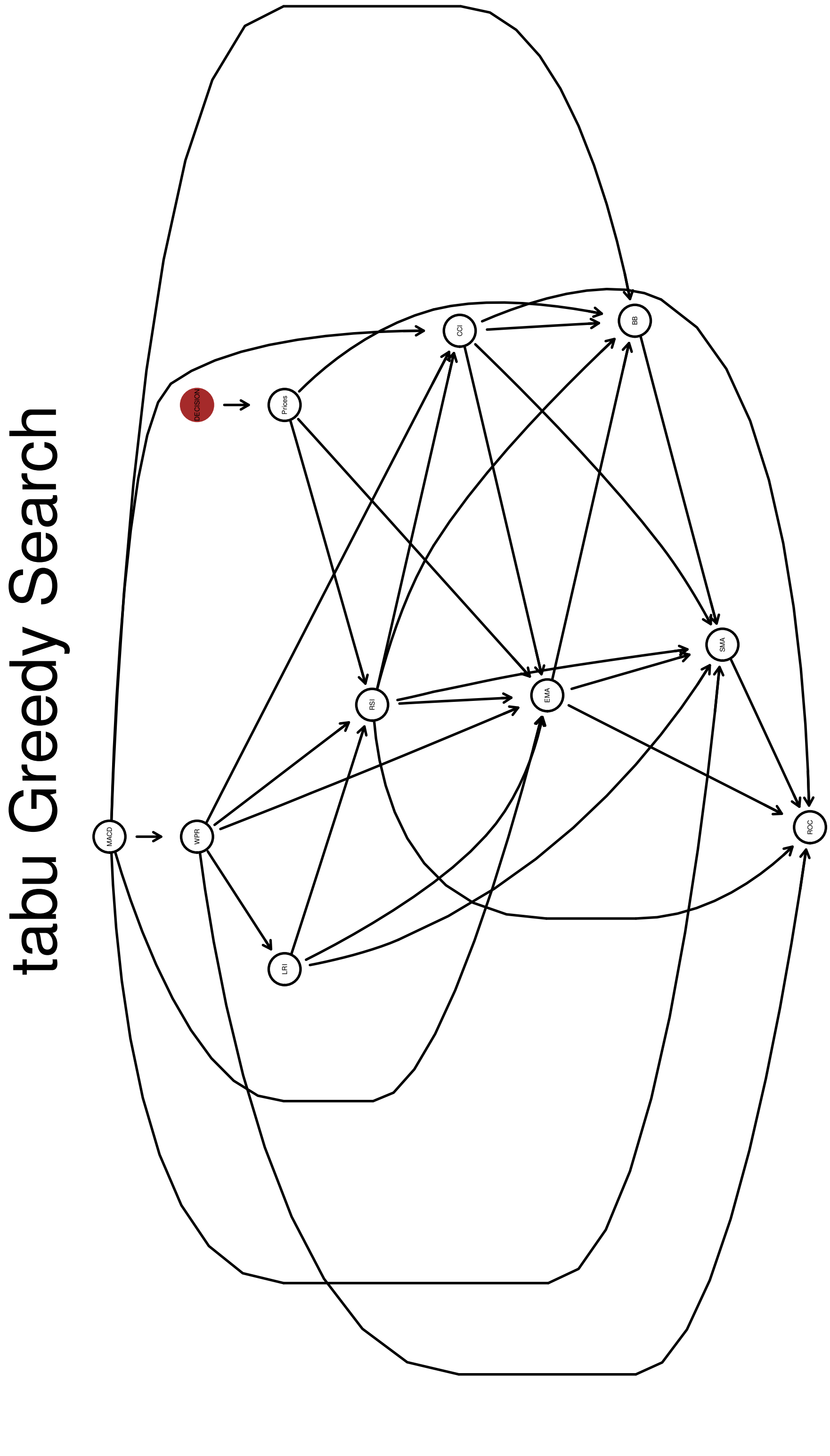


Figure 5.6: The Tabu Greedy Search BN Structure

```
Bayesian network learned via Score-based methods

model:
 [MACD][Prices][WPR|MACD][DECISION|Prices][LRI|WPR][RSI|LRI
 [CCI|MACD:WPR:RSI][EMA|LRI:MACD:WPR:CCI:RSI:Prices][SMA|LR
 [ROC|WPR:CCI:RSI:EMA:SMA][BB|MACD:CCI:SMA:Prices]
nodes:                                 11
arcs:                                  29
  undirected arcs:                     0
  directed arcs:                       29
average markov blanket size:           6.91
average neighbourhood size:            5.27
average branching factor:              2.64

learning algorithm:                    Hill-Climbing
score:                                 BIC (Gauss.)
penalization coefficient:              3.298573
tests used in the learning procedure:  405
optimized:                             TRUE
```

```
Bayesian network learned via Score-based methods

model:
 [MACD][DECISION][WPR|MACD][Prices|DECISION][LRI|WPR][RSI|LRI:WPR:Prices]
 [CCI|MACD:WPR:RSI][EMA|LRI:MACD:WPR:CCI:RSI:Prices][BB|MACD:CCI:RSI:EMA:Prices]
 [SMA|LRI:MACD:BB:CCI:RSI:EMA][ROC|WPR:CCI:RSI:EMA:SMA]
nodes:                                 11
arcs:                                  31
  undirected arcs:                     0
  directed arcs:                       31
average markov blanket size:           7.27
average neighbourhood size:            5.64
average branching factor:              2.82

learning algorithm:                    Tabu Search
score:                                 BIC (Gauss.)
penalization coefficient:              3.298573
tests used in the learning procedure:  722
optimized:                             TRUE
```

Figure 5.7: The Hill-climbing and Tabu Greedy Search learning Records

used a threshold $\alpha = 0.05$, since the constraint based algorithms are self-correcting and there is no need to use multiplicity correction to select a convenient threshold. Any threshold falling between the interval $[0.01, 0.05]$ would work just fine with networks having up to 100 variables. The number of tests used for IAMB learning are 524 and that of grow-shrink are 430.

Discretization Approaches

Here we dicretize data before learning the network to study th effect of that on the overall performance of the network. We apply various approaches based on discretized into a form that preserves the dependence structure of the data. Studying by that the change of the BN learned results. The indicator values and prices will be discretized

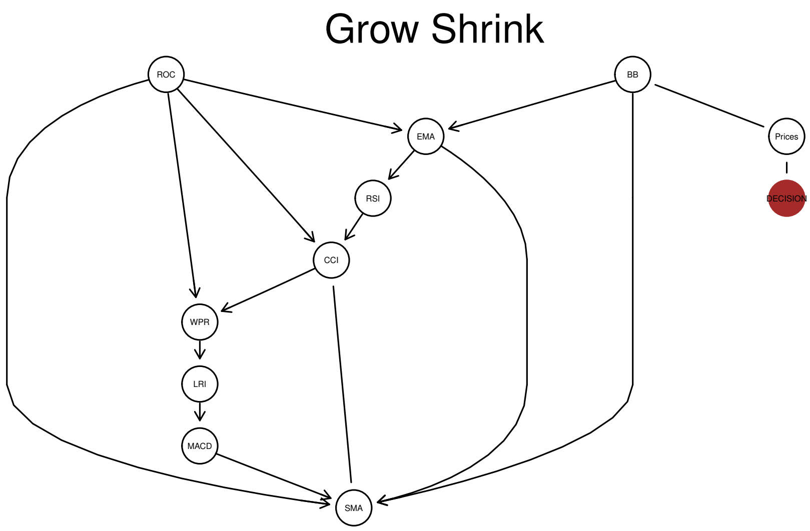


Figure 5.8: The Grow Shrink learned BN Structure

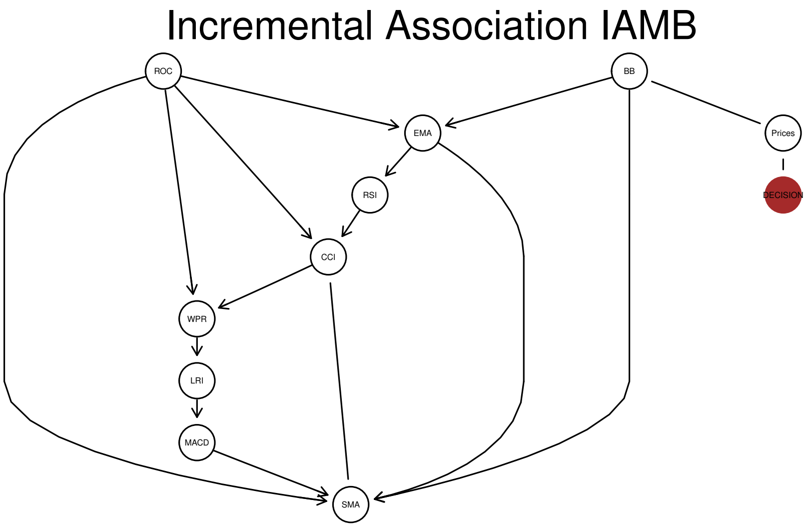


Figure 5.9: The Incremental Association learned BN Structure

```
Bayesian network learned via Constraint-based | Bayesian network learned via Constraint-based methods

model:                                          model:
  [partially directed graph]                      [partially directed graph]
nodes:                                11        nodes:                                11
arcs:                                 16        arcs:                                 16
  undirected arcs:                    3           undirected arcs:                    3
  directed arcs:                      13          directed arcs:                      13
average markov blanket size:          3.82      average markov blanket size:          3.82
average neighbourhood size:           2.91      average neighbourhood size:           2.91
average branching factor:             1.18      average branching factor:             1.18

learning algorithm:                   Grow-Sh   learning algorithm:                   IAMB
conditional independence test:        Pearson   conditional independence test:        Pearson's Correlation
alpha threshold:                      0.05      alpha threshold:                      0.05
tests used in the learning procedure: 430       tests used in the learning procedure: 524
optimized:                            TRUE      optimized:                            TRUE
```

Figure 5.10: The Grow-shrink and IAMB learning Records

into two levels interval according to the imported values from Matlab, and the decision is disretized into three levels or states representing the three possible decisions of buying, holding and selling. Another importance to discretization is that the parameters of discrete variables are conditional probability tables CPTs instead of linear regression with continuous data. Having the conditional probability tables makes it easy to test the approach for daily decisions effect on return and hit ratio. After applying the suitable discretization we can reapply any of the structure learning algorithms previously applied earlier. We choose to test with Hill climbing greedy search score based algorithm, using the same function *hc* previously introduced. Figure 5.12 illustrates the output network structure, where the decision node is directly affected by two indicators MACD and LRI. [49]. The learning record of the BN is delivered in figure 5.11, showing a full DAG being learned 605 tests scored with log-likelihood. Another Approach to apply is a setup similar to that of Sach et al [58]. It is a bootstrap re-sampling setup that allows us to learn a set of network structures (for example 500). It learns a network structure for each bootstrap sample with a pre-selected Hill-climbing structure learning algorithm and a Bayesian Dirichlet equivalent (BDe) posterior density with a low imaginary sample size. It then applies an average to the occurrence of arcs present in the network of each study taking into consideration the arc orientation. We can also choose a detection threshold to help select arcs that have shown existence in more than a certain percentage of the graphs (for example 85% of the graphs). In simple words, arcs are considered significant when they repeat to appear in 85% of the graphs in the most frequently recorded direction. An interesting fact is that the authors prove that lowering the threshold percentage to any value above than 50% does not change the outcome at all. This means, it is not worth choosing a tested value since, the latter is not a critical factor affecting the outcome. Figure 5.13 illustrates the outcome of the bootstrap approach. We can notice that this technique causes a complete independence of the decision node, where it has no kind of link with any other node in the graph. Figure 5.14 shows the report of learning to this algorithm.

Another Alternative is also applying an average but of Networks resulting from several hill climbing searches generated randomly from a uniform distribution over a space of connected graphs. This insures avoiding systematic bias covering by that the search space completely. A testing of this approach is applied with a generated learned network from its data. The Network Structure and learning Record are shown in figures 5.15 and 5.16 respectively.

```
Bayesian network learned via Score-based methods

model:
 [MACD][LRI|MACD][Decision|LRI:MACD][Prices|LRI:MACD:Decision]
 [WPR|LRI:MACD:Prices:Decision][CCI|LRI:MACD:WPR:Prices:Decision]
 [RSI|LRI:MACD:WPR:CCI:Prices:Decision][EMA|LRI:MACD:WPR:CCI:RSI:Prices:Decision]
 [SMA|LRI:MACD:WPR:RSI:EMA:Prices:Decision]
 [ROC|LRI:MACD:WPR:CCI:RSI:EMA:SMA:Prices:Decision]
 [BB|LRI:MACD:ROC:WPR:CCI:RSI:EMA:Prices:Decision]
nodes:                                 11
arcs:                                  53
  undirected arcs:                     0
  directed arcs:                       53
average markov blanket size:           9.82
average neighbourhood size:            9.64
average branching factor:              4.82

learning algorithm:                    Hill-Climbing
score:                                 Log-Likelihood (disc.)
tests used in the learning procedure:  605
optimized:                             TRUE
```

Figure 5.11: Learning Report of the Hill-climbing approach with Discretized Data

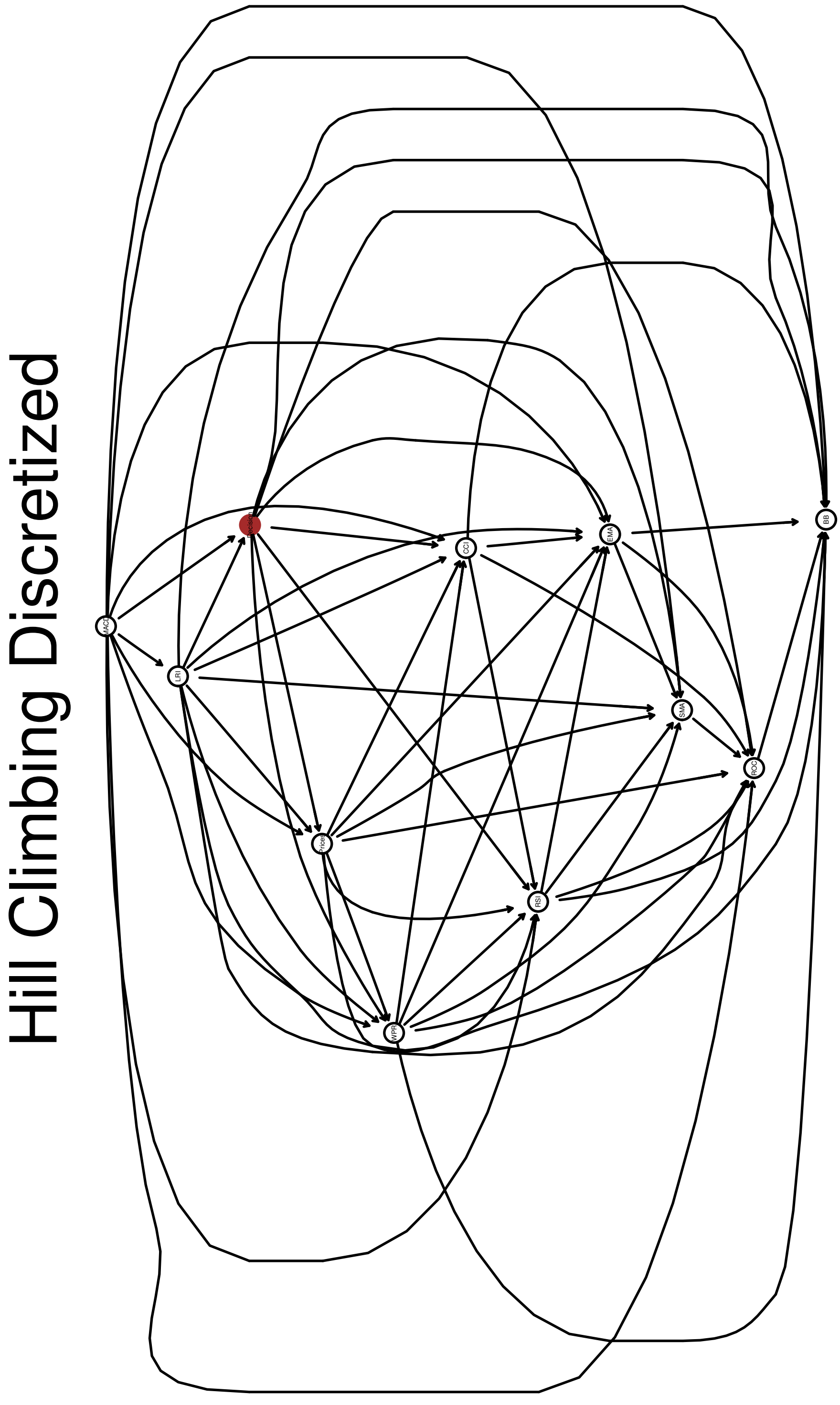


Figure 5.12: The Hill-climbing of Discretized data BN Structure

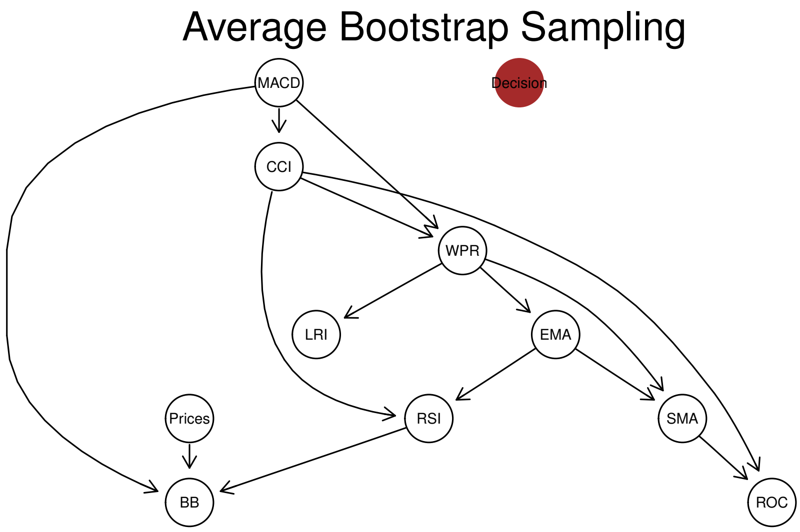


Figure 5.13: The BN structure of the Average Bootstrap Algorithm

```
Random/Generated Bayesian network

model:
 [MACD][Prices][Decision][CCI|MACD][WPR|MACD:CCI][LRI|WPR][EMA|WPR][RSI|CCI:EMA]
 [SMA|WPR:EMA][ROC|CCI:SMA][BB|MACD:RSI:Prices]
nodes:                                 11
arcs:                                  14
  undirected arcs:                     0
  directed arcs:                       14
average markov blanket size:           3.45
average neighbourhood size:            2.55
average branching factor:              1.27

generation algorithm:                  Model Averaging
significance threshold:                0.85
```

Figure 5.14: Learning Report of the Average Bootstrap Algorithm with Discretized Data

5.6.2 Parameter learning with bnlearn

The next step after learning the structure of the network, it is necessary to estimate the parameters of the local distributions. With the bnlearn package, this is applied by using the function *bn.fit*. This function takes as parameters, the data and the network structure. For continuous data the parameters take the form of regression coefficients (for more information [49]). As an illustrative example we apply the *fit* function to one of the structures learned in the basic approaches before discretization, the output of a single node is shown in figure 5.17. As for discrete data *fitted* function returnd the CPTs of nodes, refer to figure 5.18 for an illustrative example.

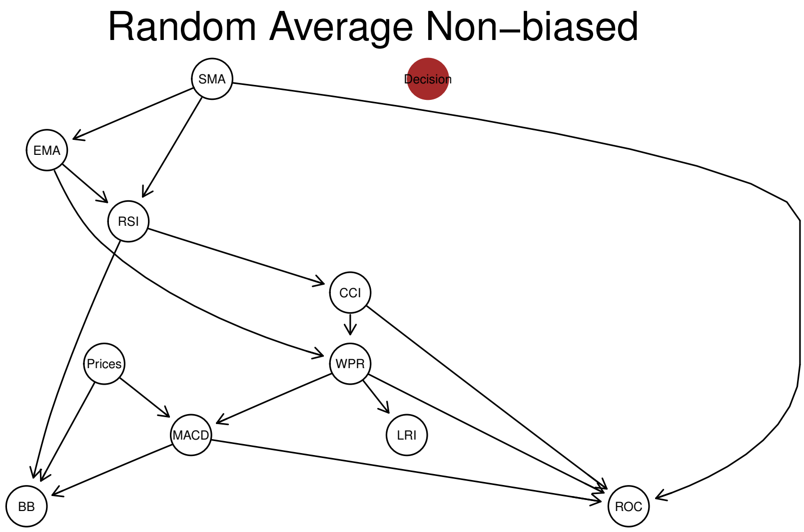


Figure 5.15: The BN structure of the Random Average Non-biased Algorithm

```
Random/Generated Bayesian network

model:
 [SMA][Prices][Decision][EMA|SMA][RSI|EMA:SMA][LRI|WPR][MACD|WPR:Prices]
 [ROC|MACD:WPR:CCI:SMA][BB|MACD:RSI:Prices][WPR|CCI:EMA][CCI|MACD:RSI]
nodes:                                 11
arcs:                                  16
  undirected arcs:                     0
  directed arcs:                       16
average markov blanket size:           4.36
average neighbourhood size:            3.09
average branching factor:              1.55

generation algorithm:                  Model Averaging
significance threshold:                0.85
```

Figure 5.16: Learning Report of the Random Average Non-biased Algorithm with Discretized Data

5.6.3 Testing the Learned Networks

In this final part, we aim to choose the network that is best efficient. In order to select the best network what is done normally is a comparison of scores. We estimate the score of each network and choose the network with the best score as our selected

```
  Parameters of node MACD (Gaussian distribution)

Conditional density: MACD
Coefficients:
(Intercept)
  0.2649601
Standard deviation of the residuals: 18.68422
```

Figure 5.17: Fitted function output for the MACD contineous Data

```
  Parameters of node Prices (multinomial distribution)

Conditional probability table:

, , MACD = [-52.7,3.57], Decision = [-1,-0.333]

                         LRI
Prices                    [-296,11.6] (11.6,319]
  [2.88e+03,4.17e+03]    0.4057971  0.3214286
  (4.17e+03,5.47e+03]    0.5942029  0.6785714

, , MACD = (3.57,59.8], Decision = [-1,-0.333]

                         LRI
Prices                    [-296,11.6] (11.6,319]
  [2.88e+03,4.17e+03]    0.4000000  0.4285714
  (4.17e+03,5.47e+03]    0.6000000  0.5714286

, , MACD = [-52.7,3.57], Decision = (-0.333,0.333]

                         LRI
Prices                    [-296,11.6] (11.6,319]
  [2.88e+03,4.17e+03]    0.4304636  0.3709677
  (4.17e+03,5.47e+03]    0.5695364  0.6290323

, , MACD = (3.57,59.8], Decision = (-0.333,0.333]

                         LRI
Prices                    [-296,11.6] (11.6,319]
  [2.88e+03,4.17e+03]    0.4000000  0.5400000
  (4.17e+03,5.47e+03]    0.6000000  0.4600000

, , MACD = [-52.7,3.57], Decision = (0.333,1]

                         LRI
Prices                    [-296,11.6] (11.6,319]
  [2.88e+03,4.17e+03]    0.4320988  0.4848485
  (4.17e+03,5.47e+03]    0.5679012  0.5151515

, , MACD = (3.57,59.8], Decision = (0.333,1]

                         LRI
Prices                    [-296,11.6] (11.6,319]
  [2.88e+03,4.17e+03]    0.5517241  0.6078431
  (4.17e+03,5.47e+03]    0.4482759  0.3921569
```

Figure 5.18: Fitted function output for the MACD Discrete Data

network for testing. However, we can notice in most of the learned graphs that the Decision node is not well correlated to the other nodes in the network (i.e. the technical indicators or the price). While, to complete the testing of the system with estimating the average return and hit ratio, we need to have a network where decision is actually affected by other nodes of the graph (indicators).

Table 5.2 shows a comparison between all learned networks. The comparison of the networks is applied according to *Bayesian information criterion* (BIC) and *Akaike Information Criterion* (AIC) scores. For more information about the scores and their defined estimation formulas refer to [49]. According to the score results we can notice that the random average non-biased approach gave the best BIC and AIC scores. Another important notice is that the grow-shrink and IAMB algorithms did not give any score results, and that is due to having non-directed arcs in these algorithms networks. They are partially directed graphs that cannot have its scores estimated.

Data Type	Learning Algorithm	BIC Score	AIC Score
Continuous Data	*Hill-climbing*	-35224.3	-35132.3
	Tabu	-35215.8	-35119.2
	Grow-shrink	N/A	N/A
	IAMB	N/A	N/A
Discrete Data	*Hill-climbing*	-10424.5	-5572.2
	Average Bootstrap	-3802.6	-3724.5
	Random Average Non-biased	-3774.8	-3664.5

Table 5.2: Comparing Scores of Different Learned BNs

In order to apply a comparison with the approaches of the hybrid probability-possibility system we use for testing the networks that have indicator nodes affecting the decision. Therefore, we use the Discretized hill climbing network that has two indicators LRI and MACD affecting its Decisions. Also it uses discretized data that allows us to use the conditional probability table of the Decision node to generate a daily decisions according to the generated value probabilities. Hence, a similar evaluation process as that introduced before 4.4.2 is applied.

The first step is estimating the conditional probability table of the Decision node, shown in figure 5.19. We can notice in the conditional probability table that the decision values is distributed into three levels, each level representing a decision.

- *Sell* interval $[-1, -0.333]$
- *Hold* interval $[-0.333, 0.333]$
- *Buy* interval $[0.333, 1]$

We selected the interval where each decision has the highest probability. The green represents buying, red represents selling and black represents holding. Afterwards, we estimate according to the indicator values falling into the selected interval, the decision to take. Note that the only interfering indicators here are MACD and LRI. Table 5.3 shows the results of the Bayesian approach along with the probability possibility hybrid approaches with respect to Average return percent and Hit ratio.

```
Parameters of node Decision (multinomial distribution)

Conditional probability table:

, , MACD = [-52.7,3.57]

                    LRI
Decision            [-296,11.6] (11.6,319]
  [-1,-0.333]         0.2292359   0.2276423
  (-0.333,0.333]      0.5016611   0.5040650
  (0.333,1]           0.2691030   0.2682927

, , MACD = (3.57,59.8]

                    LRI
Decision            [-296,11.6] (11.6,319]
  [-1,-0.333]         0.2293578   0.2450000
  (-0.333,0.333]      0.5045872   0.5000000
  (0.333,1]           0.2660550   0.2550000
```

Figure 5.19: Conditional Probability Tabel CPT of Decision Node in Discretized Hill-climbing Algorithm

Individual Indicators vs. Fusion Techniques		**EURO STOXX 50 %AR**	**%HR**
ROC		0.3110	69.1
LRI		0.6942	67.4
MACD		1.2491	76.7
BB		0.4941	41
Majority Vote		1.067	77.2
Non-weighted Possibility Fusion	*MoMaxs*	0.3726	58.1
	MoAvgs	1.5026	85.4
	MoMins	0.8834	72.7
Weighted Possibility Fusion	*MoMaxs*	0.3441	57.1
	MoAvgs	2.0778	82.5
	MoMins	0.5575	63.9
Dynamic Weighted Possibility Fusion	*MoMaxs*	1.3251	68.5
	MoAvgs	3.0034	85.7
	MoMins	1.9240	67.5
Bayesian Network Fusion		0.7592	64.5

Table 5.3: Comparing Fusion Approaches with BN Approach with respect to hit Ratio and average return

We can notice in table 5.3 that Bayesian fusion did not overcome the best performing hybrid approaches or Individual Indicator. We have mentioned that contributing indicators to the BN approach are MACD and LRI, which showed better performance

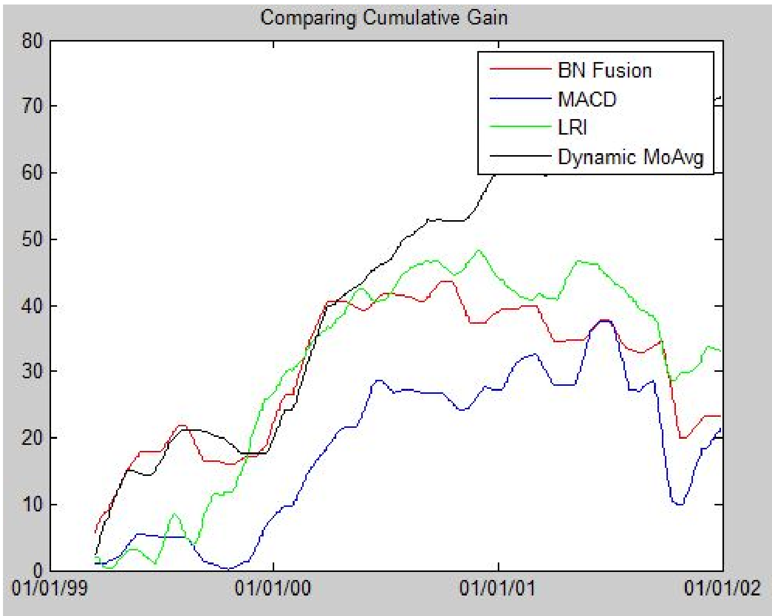


Figure 5.20: Cumulative Gain Comparison Plot

individually than the Bayesian fusion approach. Also a plot of cumulative gain is shown in figure 5.20. The plot compares the performance of MACD and LRI individual indicators with dynamic possibility fusion MoAvgs approach and BN fusion approach with respect to cumulative gain. We can see how the dynamic MoAvgs overcame all other approaches. However, for the sake of transparency, the Bayesian Network approach in this work was used in a very basic manner for comparison purposes only. Very few tests were performed and the approach could be optimized for better results. Another reason to mention is that the studied network is not the best scored one for the problem of lack of connection with the Decision node.

5.7 Conclusion

This chapter introduced Bayesian Networks in details with its tools and for network learning and decision making. We used it to introduce a purely probabilistic fusion approach to compare to the proposed contribution of probability possibility approaches discussed in the previous chapter. A basic learning and testing was applied to the indicators fusion problem, using the *bnlearn* package functions in R. We have learned various networks and noticed that most of them have no influence of indicators on the decision. This forced us to apply the testing on the only available network exhibiting an influence of indicators on the decision, although it was not the best scoring network. Therefore, we cannot judge the efficiency of using Bayesian Networks with indicators fusion for the reason of lack of efficient testing. As for the results we content ourselves by considering this approach a basis for future optimization and potential success.

CHAPTER 6 Conclusion

After forming a clear understanding of the situation challenges and contributions, it becomes necessary to summarize all derived conclusions and consequences with a declaration of a fair personal perspective on the applied studies.

The subject of this thesis was aroused by the undoubted need of investors to have a robust and risk discounted decision support system that recommends the right winning decision at the right time. Studying available theories concerned with the target goal, oriented our attraction towards technical analysis techniques and its tools, in particular technical indicators. This type of analysis is considered very interesting to this research objective. However, it is evident that the whole situation is integrated with uncertainty, human emotions, and ambiguity. This led into an intuitive need to dig into different artificial intelligence reasoning methods, that best deal with the challenges of this research.

The first applied approach is the hybrid probability-possibility approach. It includes four proposed decision support systems that takes advantage of the powers of both probability and possibility theories overcoming by that the limitations of each theory when deployed solely. Chapter 4 includes an expansive and precise description of the developed decision support system along with an exhaustive testing plan tight enough to strain all used parameters and affecting factors for a transparent and fair judgment of the contributed solutions.

A brief look into the results suggests that, some decision support systems showed better performance than expected. The complete scheme of testing should be enough to give a clear perception about which parameters and indicators to choose for different analysts. Furthermore, the testing applied covered both in-sample and out-of-sample testing, since the testing was applied not only on different data period than the one used for learning and optimization, but also on a totally different index (CAC40). It is also very interesting that most fusion approaches induced gain, even though the less innovative techniques where overcame by other more efficient techniques, like weighted and dynamic weighted fusion techniques. Another important thing to mark is the very inevitable additive effect of including dynamic reliability factors to the fusion (with MoAvg dynamic fusion technique being the most successful), and the significant accomplishment of weighted MoMin when used with (90 days, 4%) winning dates estimation parameters on both scales of Average return and Hit Ratio.

Chapter 5 introduced an additional basic fusion approach based on pure probability-based analysis with Bayesian Networks to be compared to the contribution of this thesis (hybrid probability possibility approach). A basic learning and testing was applied to the indicators fusion problem, using the *bnlearn* package functions in R. Various networks have been learned, with the observation of most having no influence of indicators on the decision. For that purpose, the testing process was not applied on the best scoring network. The testing was rather applied on the only available network that showed an influence of indicators on the decision. Therefore, in our opinion it is not fairly accurate to judge the efficiency/inefficiency of using Bayesian Networks with indicators fusion for the reason of lack of sufficient testing and regulation. As for the results we content ourselves by considering this approach a basis for future optimization and potential success.

As a personal perspective, we can confidently certify that the applied research has met the objective of this thesis. The main challenge of superiority of merged indicators over individual one was achieved in many of the proposed decision support systems. Furthermore, the inclusion of reliability and its dynamic form is an innovative contribution added to the current available body of knowledge. This proposed work plan is just the door to many potential financial decision support systems. An example of possible extensions, would be an integration of the system with a Markowitz portfolio allocation system (discussed in chapter 1, section 1.7.1) that makes use of the decision output of the hybrid probability possibility decision fusion approach. A quick description to the logic of the picking process, the system checks every γ days the highest degree of confidence for a decision of a collection of securities. According to that degree the system changes the weight of securities in the portfolio (buying a security when it records a high degree of buying, and selling it when it records a high degree of selling derived from the applied hybrid systems) with making sure of keeping a constant investment value. Finally, we believe that the field of reasoning methods application in finance could be considered as a mine of research, where lies an unlimited horizon of innovation.

List of Publications

1. A. Itani, JM. Le Caillec, B. Solaiman and A. Hamié. Hybrid Probability-Possibility Decision Support Systems for Merging Technical Indices. 17^{th} *International Conference on Information Fusion*, July 2014 (*Fusion 2014, Salamanca, Spain*).

2. A. Itani, JM. Le Caillec, B. Solaiman and A. Hamié. Probability-Possibility Hybrid Systems for Merging Technical Indices. Submitted, *Journal Traitement du Signal Lavoisier*, September 2013.

3. A. Itani, JM. Le Caillec, B. Solaiman, A. Hamié. Fuzzy Approach for Merging Technical Indices. *Traitement et Analyse de l'information Méthodes et Applications*, May 2013 (*TAIMA 2013, Hammamet, Tunisia*).

Bibliography

[1] Ajith Abraham and Baikunth Nath. Hybrid intelligent systems design: A review of a decade of research. *IEEE Trans-actions on Systems, Man and Cybernetics (Part C)*, 3(1):1–37, 2000. 60

[2] Ajith Abraham, Baikunth Nath, and Prabhat Kumar Mahanti. Hybrid intelligent systems for stock market analysis. In *Computational Science-ICCS 2001*, pages 337–345. Springer, 2001. 60

[3] Steven B Achelis. *Technical Analysis from A to Z*. McGraw Hill New York, 2001. 34

[4] Amit Agarwal. Technical analysis. Technical report, Ben Graham School of Finance, 2013. 35

[5] Franklin Allen and Risto Karjalainen. Using genetic algorithms to find technical trading rules. *Journal of financial Economics*, 51(2):245–271, 1999. 59

[6] George Atsalakis, Camelia Ucenic, and C Skiadas. Time series prediction of the greek manufacturing index for the non-metallic minerals sector using a neuro-fuzzy approach (anfis). In *International symposium on applied stochastic models and data analysis, France, Brest*, 2005. 60

[7] G Ramesh Babu. *Portfolio Management (including Security Analysis)*. Concept Publishing Company, 2007. 8

[8] Jørgen Bang-Jensen and Gregory Gutin. Theory, algorithms and applications. *Springer Monographs in Mathematics, Springer-Verlag London Ltd., London*, 2007. 101

[9] John Bollinger. *Bollinger on Bollinger bands*. McGraw Hill Professional, 2001. 48

[10] Eike Christain Brechmann and Claudia Czado. Risk management with high-dimensional vine copulas: An analysis of the euro stoxx 50. *Statistics & Risk Modeling*, 30(4):307–342, 2013. 86

[11] William Brock, Josef Lakonishok, and Blake LeBaron. Simple technical trading rules and the stochastic properties of stock returns. *The Journal of Finance*, 47(5):1731–1764, 1992. 57

[12] Niko Canner, N Gregory Mankiw, and David N Weil. An asset allocation puzzle. Technical report, National Bureau of Economic Research, 1994. 19

[13] Gunther Capelle-Blancard and Mo Chaudhury. Efficiency tests of the french index (cac 40) options market. *McGill Finance Research Centre Working Paper*, 2001. 86

[14] Christer Carlsson and Robert Fullér. A neuro-fuzzy system for portfolio evaluation. *Cybernetics and Systems Research*, pages 296–299, 1996. 60

[15] Wee Mien Cheung and Uzay Kaymak. A fuzzy logic based trading system. In *Proceedings of the Third European Symposium on Nature-inspired Smart Information Systems*, 2007. 59

[16] Ta-Chung Chu, Chung-Tsen Tsao, and Yeou-Ren Shiue. Application of fuzzy multiple attribute decision making on company analysis for stock selection. In *Fuzzy Systems Symposium, 1996. Soft Computing in Intelligent Systems and Information Processing., Proceedings of the 1996 Asian*, pages 509–514. IEEE, 1996. 59

[17] Wikipedia contributors. Information ratio, 2013. 24

[18] Wikipedia contributors. Beysian networks, 2014. 112

[19] Wikipedia contributors. Graph theory, 2014. 100

[20] Wikipedia contributors. Information theory, 2014. 79

[21] Wikipedia contributors. Treynor ratio, 2014. 23

[22] Chad Langager Cory Janssen and Casey Murphy. Technical analysis: Chart types, 2013. 37

[23] Thomas M Cover and Joy A Thomas. Entropy, relative entropy and mutual information. *Elements of Information Theory*, pages 12–49, 1991. 79, 80

[24] Michael Alan Howarth Dempster and Chris M Jones. *The profitability of intra-day FX trading using technical indicators*. Judge Institute of Management, University of Cambridge, 2000. 34

[25] Claus Dethlefsen and Susanne G Boettcher. deal: A package for learning bayesian networks. *Journal of Statistical Software*, 8(20), 2003. 114

[26] Reinhard Diestel. Graph theory. 2005. *Grad. Texts in Math*, 2005. 101

[27] Hussein Dourra and Pepe Siy. Investment using technical analysis and fuzzy logic. *Fuzzy sets and systems*, 127(2):221–240, 2002. 59

[28] Didier Dubois, Laurent Foulloy, Gilles Mauris, and Henri Prade. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable computing*, 10(4):273–297, 2004. 73, 74

[29] Didier Dubois and Henri Prade. Possibility theory in information fusion. In *Information Fusion, 2000. FUSION 2000. Proceedings of the Third International Conference on*, volume 1, pages PS6–P19. IEEE, 2000. 63, 64

[30] Didier Dubois and Henri Prade. Possibility theory and its applications: Where do we stand? *Mathware and Soft Computing*, 18(1):18–31, 2011. 61, 62

[31] Didier J Dubois. *Fuzzy sets and systems: theory and applications*, volume 144. Academic press, 1980. 61, 73

[32] Edwin J Elton, Martin J Gruber, Stephen J Brown, and William N Goetzmann. *Modern portfolio theory and investment analysis*. John Wiley & Sons, 2009. 9, 19

[33] Fernando Fernández-Rodríguez, Christian González-Martel, and Simón Sosvilla-Rivero*. Optimization of technical rules by genetic algorithms: evidence from the madrid stock market. *Applied Financial Economics*, 15(11):773–775, 2005. 59

[34] Joseph E Granville. *Granvilleś New Key to Stock Market: Profits*. Prentice-Hall, 1968. 48

[35] Sally Hamilton, Hoje Jo, and Meir Statman. Doing well while doing good? the investment performance of socially responsible mutual funds. *Financial Analysts Journal*, pages 62–66, 1993. 23

[36] Ypke Hiemstra. A stock market forecasting support system based on fuzzy logic. In *System Sciences, 1994. Proceedings of the Twenty-Seventh Hawaii International Conference on*, volume 3, pages 281–287. IEEE, 1994. 58

[37] Westpac Online Investing. The basics of fundamental analysis. Technical report, 2010. 29

[38] Michael C Jensen, Fischer Black, and Myron S Scholes. The capital asset pricing model: Some empirical tests. 1972. 19, 20

[39] Markus Kalisch and Peter Bühlmann. Estimating high-dimensional directed acyclic graphs with the pc-algorithm. *The Journal of Machine Learning Research*, 8:613–636, 2007. 114

[40] Myoung Jong Kim, Ingoo Han, and Kun Chang Lee. Hybrid knowledge integration using the fuzzy genetic algorithm: prediction of the korea stock price index. *Intelligent Systems in Accounting, Finance and Management*, 12(1):43–60, 2004. 60

[41] George J Klir and Behzad Parviz. Probability-possibility transformations: a comparison. *International Journal of General System*, 21(3):291–310, 1992. 73

[42] Kevin B Korb and Ann E Nicholson. *Bayesian artificial intelligence*. cRc Press, 2003. 103, 106

[43] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, pages 79–86, 1951. 82

[44] Donald R Lambert. Commodity channel index: Tool for trading cyclic trends. *Technical Analysis of Stocks & Commodities*, 1, 1983. 47

[45] William Leigh, Russell Purvis, and James M Ragusa. Forecasting the nyse composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in< i> romantic</i> decision support. *Decision Support Systems*, 32(4):361–377, 2002. 57

[46] Robert A Levy. The predictive significance of five-point chart patterns. *Journal of Business*, pages 316–323, 1971. 57

[47] Chin-Shien Lin, Haider A Khan, and Chi-Chung Huang. Can the neuro fuzzy model predict stock indexes better than its rivals. *Discussion Papers of University of Tokyo CIRJE-F-165*, 2002. 60

[48] Andrew W Lo and Archie Craig MacKinlay. Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of financial studies*, 1(1):41–66, 1988. 34

[49] Radhakrishnan Nagarajan, Marco Scutari, and Sophie Lèbre. *Bayesian Networks in R*. Springer, 2013. 122, 125, 127

[50] Christopher Neely, Paul Weller, and Rob Dittmar. Is technical analysis in the foreign exchange market profitable? a genetic programming approach. *Journal of financial and Quantitative Analysis*, 32(04):405–426, 1997. 34

[51] CV Negoita, LA Zadeh, and HJ Zimmermann. Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 1:3–28, 1978. 61, 72

[52] Iyiola Omisore, Munirat Yusuf, and Nwufo Christopher. The modern portfolio theory as an investment decision tool. *Journal of Accounting and Taxation*, 4(2):19–28, 2011. 18

[53] Mourad Oussalah. On the probability/possibility transformations: a comparative analysis. *International Journal Of General System*, 29(5):671–718, 2000. 73

[54] Konstantinos N Pantazopoulos, Lefteri H Tsoukalas, Nikolaos G Bourbakis, Michael J Brun, and Elias N Houstis. Financial prediction and trading strategies using neurofuzzy approaches. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 28(4):520–531, 1998. 60

[55] Judea Pearl, Thomas Verma, et al. *A theory of inferred causation*. Morgan Kaufmann San Mateo, CA, 1991. 113, 114

[56] Frank Reilly and Keith Brown. *Investment analysis and portfolio management*. Cengage Learning, 2011. 7

[57] Stuart Russell. Artificial intelligence: A modern approach author: Stuart russell, peter norvig, publisher: Prentice hall pa. 2009. 113

[58] Karen Sachs, Omar Perez, Dana Pe'er, Douglas A Lauffenburger, and Garry P Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005. 123

[59] Marco Scutari. Learning bayesian networks with the bnlearn r package. *arXiv preprint arXiv:0908.3817*, 2009. 118

[60] Marco Scutari. Measures of variability for graphical models. 2011. 113

[61] William F Sharpe. The sharpe ratio. *Streetwise–the Best of the Journal of Portfolio Management*, pages 169–185, 1998. 23

[62] Rimvydas Simutis. Fuzzy logic based stock trading system. In *Computational Intelligence for Financial Engineering, 2000.(CIFEr) Proceedings of the IEEE/IAFE/INFORMS 2000 Conference on*, pages 19–21. IEEE, 2000. 59

[63] Sameer Singh and Jonathan Fieldsend. Financial time series forecasts using fuzzy and long memory pattern recognition systems. In *Computational Intelligence for Financial Engineering, 2000.(CIFEr) Proceedings of the IEEE/IAFE/INFORMS 2000 Conference on*, pages 166–169. IEEE, 2000. 59

[64] Ezra Solomon et al. Return on investment: The relation of book-yield to true yield. In *Fall Meeting of the Society of Petroleum Engineers of AIME*. Society of Petroleum Engineers, 1963. 86

[65] Hia Jong Teoh, Ching-Hsue Cheng, Hsing-Hui Chu, and Jr-Shian Chen. Fuzzy time series model based on probabilistic approach and rough set rule induction for empirical research in stock markets. *Data & Knowledge Engineering*, 67(1):103–117, 2008. 60

[66] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131, 1974. 88

[67] Jan van den Berg, Uzay Kaymak, and Willem-Max van den Bergh. Financial markets analysis by using a probabilistic fuzzy modelling approach. *International Journal of Approximate Reasoning*, 35(3):291–305, 2004. 60

[68] J Wells Wilder Jr. The relative strength index ,Äú. *J. of Technical Analysis of Stocks and Commodities*, 1986. 44, 45

[69] Du Xu and Uzay Kaymak. Value-at-risk estimation by using probabilistic fuzzy systems. In *Fuzzy Systems, 2008. FUZZ-IEEE 2008.(IEEE World Congress on Computational Intelligence). IEEE International Conference on*, pages 2109–2116. IEEE, 2008. 60

[70] Xu-Shen Zhou and Ming Dong. Can fuzzy logic make technical analysis 20/20? *Financial Analysts Journal*, pages 54–75, 2004. 57

List of Figures

List of Tables

Résumé

La gestion de portefeuille en finance consiste à faire des bénéfices tout en minimisant le risque. Cependant, la difficulté principale de cette opération réside dans la nature volatile des prix des titres sur le marché. Pour cette raison, des techniques d'analyse de titres ont été développées pour aider les gérants de portefeuille à prévoir les changements futurs des prix afin de prendre des décisions pertinentes.
La première technique est l'analyse fondamentale qui est basée sur une étude détaillée des facteurs fondamentaux de l'entreprise émettant le titre. Cette analyse est complexe, consommatrice de temps et sujette à certaines réticences après le développement de la finance comportementale qui a remis en cause les notions d'efficacité des marchés. Cette remise en cause a suscité l'intérêt des gestionnaires de portefeuille pour un autre type d'analyse, l'analyse technique qui est le point de départ du travail de thèse. L'examen des forces et des faiblesses de chaque approche suggère l'application d'une approche hybride profitant des théories des probabilités et des possibilités.
Notre démarche a été de montrer que la fusion de plusieurs indicateurs techniques peut conduire à de meilleures décisions que celles basées sur un indicateur seul afin de prédire les variations de prix et de tendance et donc de prendre une décision d'achat ou de vente pertinente. Nous avons proposé des systèmes de décisions hybrides pour effectuer la fusion et nous avons incorporé un coefficient prenant en compte la fiabilité des indicateurs dans le processus de fusion.
Les systèmes définis ont été testés de manière exhaustive, transparente et ont montré des résultats prometteurs.

Mots-clés : Théorie de Probabilité, Théorie de Possibilité, Fusion d'information, Réseaux Bayesiens, Théorie d'information, Divergence de Kullback-Leibler.

Abstract

Portfolio management is a mean of making profit and expanding wealth through following different security trading strategies, such as the act of buying a financial security at a certain price, and selling it later at a higher price to make profit out of this trade. However, the main difficulty lies in the diligently varying nature of security prices in the financial market.
In this research we apply multiple decision support systems to help investors and traders in making the right decision at the right time. These systems belong to a hybrid approach that takes advantage of both theories of probability and possibility. Since, probability theory is known for its power with learning statistical data, and possibility theory is known for its competences in handling uncertainty and processing any human factor incorporation. Thus, perfectly handles the challenges included in the situation understudy.
The main objective of this applied research is studying the effects of multiple Technical indicators fusion on the risk and revenue upon making a trading decision. By that, taking advantage of the Kullback Leibler divergence and Dubois-Prade transformation techniques to provide each indicator with a weight factor that represents its efficiency.
The work is completed by an exhaustive and transparent testing, comparison and validation of all the developed systems.

Keywords : Probability Theory, Possibility Theory, Information Fusion, Bayesian Networks, Information Theory, Kullback Leibler Divergence.

TELECOM Bretagne

n° d'ordre : 2014telb0348

Télécom Bretagne
Technopôle Brest-Iroise - CS 83818 - 29238 Brest Cedex 3
Tél : + 33(0) 29 00 11 11 - Fax : + 33(0) 29 00 10 00