St. Petersburg University
Graduate School of Management

Master in Management Program

IBM Watson Analytics vs. Conventional Econometrical Software: A Comparative Analysis of Suitability for Financial Sector

Master's Thesis by the 2nd year student

Concentration — Information Technologies and Innovative Management

Ilias Faizullov

Research advisor:

Associate professor, Sergey A. Yablonsky
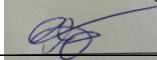
St. Petersburg

2016

ЗАЯВЛЕНИЕ О САМОСТОЯТЕЛЬНОМ ХАРАКТЕРЕ ВЫПОЛНЕНИЯ ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

Я , Файзуллов Ильяс Рафисович, студент второго курса магистратуры направления «Менеджмент», заявляю, что в моей магистерской диссертации на тему «IBM Watson Analytics и Стандартные Эконометрические Пакеты: Сравнительный анализ пригодности для финансового сектора», представленной в службу обеспечения программ магистратуры для последующей передачи в государственную аттестационную комиссию для публичной защиты, не содержится элементов плагиата.

Все прямые заимствования из печатных и электронных источников, а также из защищенных ранее выпускных квалификационных работ, кандидатских и докторских диссертаций имеют соответствующие ссылки.

Мне известно содержание п. 9.7.1 Правил обучения по основным образовательным программам высшего и среднего профессионального образования в СПбГУ о том, что «ВКР выполняется индивидуально каждым студентом под руководством назначенного ему научного руководителя», и п. 51 Устава федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет» о том, что «студент подлежит отчислению из Санкт-Петербургского университета за представление курсовой или выпускной квалификационной работы, выполненной другим лицом (лицами)».
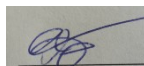
_____ (Подпись студента)

25.05.2016 (Дата)

STATEMENT ABOUT THE INDEPENDENT CHARACTER OF THE MASTER THESIS

I, Faizullov Ilias, (second) year master student, program «Management», state that my master thesis on the topic « IBM Watson Analytics vs. Conventional Econometrical Software: A Comparative Analysis of Suitability for Financial Sector », which is presented to the Master Office to be submitted to the Official Defense Committee for the public defense, does not contain any elements of plagiarism.

All direct borrowings from printed and electronic sources, as well as from master theses, PhD and doctorate theses which were defended earlier, have appropriate references.

I am aware that according to paragraph 9.7.1. of Guidelines for instruction in major curriculum programs of higher and secondary professional education at St.Petersburg University «A master thesis must be completed by each of the degree candidates individually under the supervision of his or her advisor», and according to paragraph 51 of Charter of the Federal State Institution of Higher Education Saint-Petersburg State University «a student can be expelled from St.Petersburg University for submitting of the course or graduation qualification work developed by other person (persons)».

_____ (Student's signature)

25.05.2016 (Date)

Table of content

Introduction

Market of financial analytics is a fast growing niche. It was estimated in Financial Analytics Market forecast conducted by Research and Markets (2014), that by 2018, total market value of financial analytics will reach the mark of 6.65 billion dollars. At the moment, many of different players are struggling to capture a share of the market, among them are such a renowned giants as IBM and Microsoft.

Such a rapid growth of the financial analytics market is driven by the need of financial organizations to manage increasing amounts of structured and unstructured information coming from different sources, as it states Srivastava (2015). In other words, emergence of big data creates a market for advanced analytics.

One of the spheres of financial analysis, which attracts attention of both financial organizations and individual traders, is stock price forecasting. Main characteristic of any financial assets, which is available for all participants of the market, is its price. These prices can be represented as prices of purchase of bonds and stocks, as currency exchange rates, or as interest rates of a bank deposit. The whole assembly of all these values at any given moment in time comprises the conjuncture of the market. There are three main classic methods of the stock price's dynamic prediction: Technical Analysis, Fundamental Analysis, and Quantitative Analysis.

According to Schwager (1996), Technical analysis is based on the examination of historical trends on the market, which are represented by the market statistic of stock prices and volumes. Technical analysis operates under the assumption that all available and relevant information, including so-called fundamental factors is reflected in the asset's prices. In addition, technical analyst assumes that some patterns of the stock market are repetitive and can be revealed using indicators, oscillators, and other "technical" methods. The shortcoming of such a methods is an absence of a systematic and scientific basing of the majority of its empirical methods.

Another approach is the Fundamental Analysis. It is based on the evaluation of the fundamental macroeconomic and microeconomic factors. Niemira (1998) claims that, fundamental analysis focuses on the condition of the issuer, on its revenues, market position, etc. Macroeconomic factors, influencing the whole industry and the country (GDP, Unemployment rates, and so on), are also taken into consideration.

The third classic approach to the stock market analysis, as it was described by Curthberston (1996), is Quantitative Analysis, which is based on statistical data, just like technical analysis, but instead of indicators, it uses statistical and mathematical models and tools, which are also referred as econometrical.

A new approach has emerged in the last years – the predictive analytics. It has gained attention due to the increasing amount of the available and relevant to the market information. Mark E. (2006) has estimated, that in 2007-2009 the humanity has generated more information than in the previous 1000 years. This information overload caused the emergence of the term "Big Data", which refers to the high-volume, high-velocity, and high-variety data. Predictive analytics is a quantitative analysis per se, but with the ability to use it on the "big data". It uses the same statistical and mathematical tools as quantitative analysis, however, it differs in the research approach: while standard econometric models just test pre-generated, based on theory hypothesis, predictive analytics is capable of finding correlations between variables in huge datasets without preliminary hypothesis i.e. predictive analytics generates its own statistical hypothesis based on the data.

Big Data creates challenges as well as opportunities, financial organization, such as banks have a lot to gain from analyzing Big Data, as Tian (2015, 34) argues: "The large scale of data contain enormously valuable information, and analytics based on big data can provide financial organizations with more business opportunities and the possibility to gain a more holistic view of both market and customers. Big data analytics can benefit banking and financial market firms in many aspects, such as accurate customer analytics, risk analysis and fraud detection. These approaches can lead to smarter and more intelligent trading, which can help organizations to avoid latent risks and provide more personalized services, thus to get a higher degree of competition advantage". Challenges of analyzing vast amount of high volume, high velocity, and high variety data, which is also presented in both unstructured and structured form, create the need for an advanced analytical tool.

Nowadays, there are multiple analytical platforms available for banking and other financial organizations. Such giants as IBM, Microsoft, Google, and Amazon are offering their analytical products to the market. According to Gartner's Magic Quadrant of Advance Analytical Platforms (2014), the leading position on the market of analytics platforms belongs to the IBM Corporation, RapidMiner, and SAS. Such a giant as Microsoft is lagging behind, but in a past two years it has showed positive dynamics and now it is catching up with the leaders.

The goal of this research is to determine which of these analytical platforms fits better for fit for the purposes of stock market forecasting. In the theoretical part, we will discuss the influence of the big data and predictive analytics on financial organization's operations. Then we will define the requirements of these organizations to an analytical platform, and generate the set of KPIs to evaluate the platforms.

Among other KPI's we will pay attention to the ability of analytical platforms (Using IBM Watson Analytics as an example) to generate predictive models for stock prices forecasting.

We will compare the results with the outcomes of some of the traditional, theoretically based econometric models.

Chapter I. The state of art predictive analytics.

1.1 Predictive analytics and big data.

Predictive analytics is connected with the term "Big Data", which has become popular in the past decade as it shown by Jianzheng (2016); figure 1 illustrates the raising academic interest to the subject.
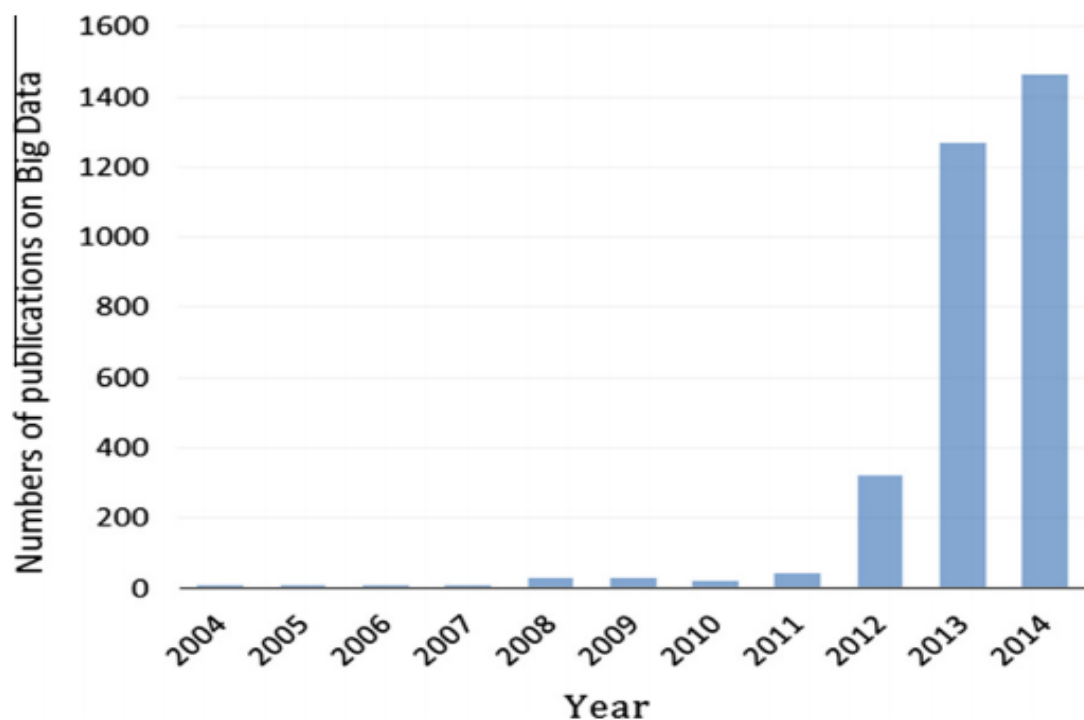
Figure 1. Dynamics of the number of published studies on Big Data. Source: Jianzheng (2016).

There is a confusion among executives around the world, regarding the question what Big Data really is. As it is shown on the figure 2, according to research conducted by SAP (2012), the majority of executives perceive big data as an increased amount of customer related information, which requires processing (28% of respondents), and almost a quarter connects Big Data with the technologies for processing vast amounts of information.

TechAmerica Foundation defines big data as follows: "Big data is a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information."

Another definition of the big data we can find at the Gartner IT Glossary: "Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."
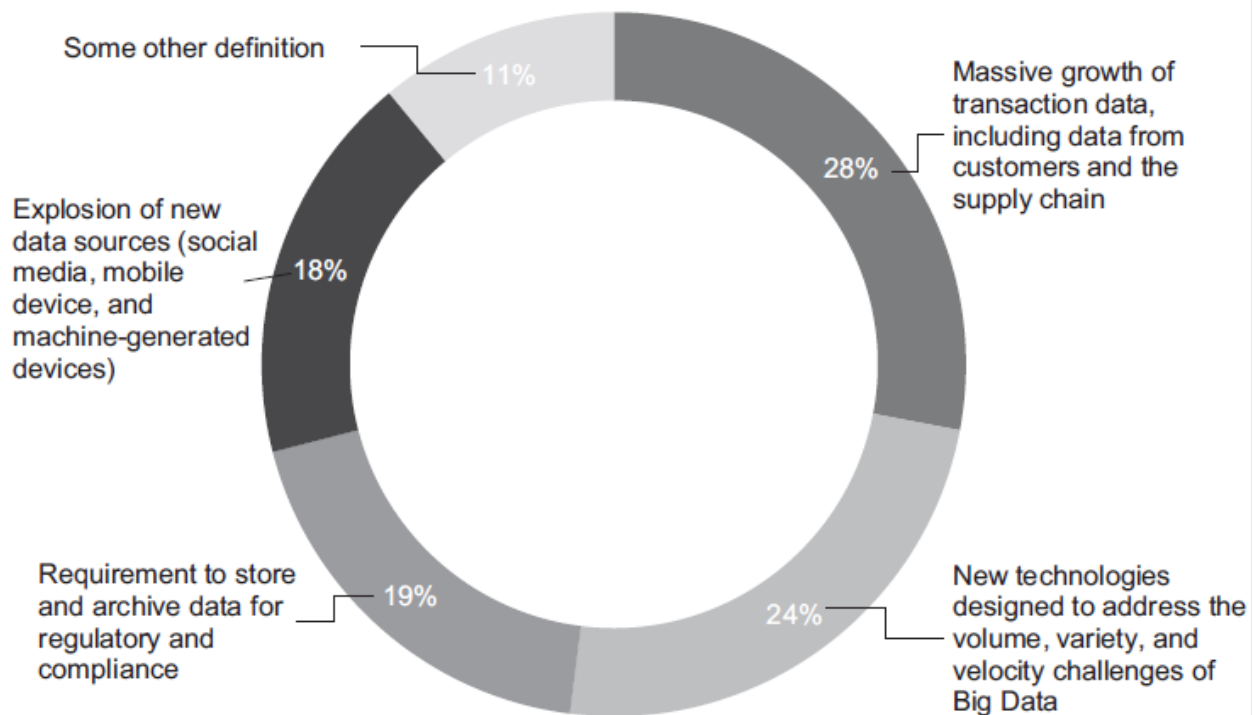
Figure 2. Definitions of big data based on an online survey of 154 executives in April 2012. Source: SAP (2012)

Both definitions describe the Big Data as a data, which possess three qualities, or, as they are also called, three V's: Volume, Variety, and Velocity.

Volume is a relative characteristic of the Big Data, as it tends to increase over time: what is considered as huge volume today may not meet the requirements of being "Big" in the future, for example ,in 2012 a dataset over a terabyte was considered as a Big Data, says Schroeck, M (2012).

Variety of the data means structural heterogeneity of the Big Data, which consist of many data formats. As Cukier, K (2010) claims, only around 5% of the data is structured, other 95 % are unstructured and represented mostly by audio, video, and text formats. Unstructured dada can not be analyzed by the machinery, therefore it poses serious challenge for an analysist.

Velocity refers to the speed at which the data is generated. The rise of the digital technologies has led to the increase of the information generation rate, making the analysis of the market even more complicated.

There are another V's of the Big Data, introduced by the IBM, SAS, and Oracle: Veracity, Variability and Complexity, and Value.

Veracity refers to the unreliability of the data, for example, social media sources are unreliable by nature, as they are generated by the broad masses of people.

Variability and Complexity refers to the unsteady rate of information generation and the diversity of the sources it comes from. Analyzing multiple information flows, which are coming at the different rates and have their own cycles, downs, and peaks, drives the need for the advanced analytics tools.

Finally, the last V – Value. The low share of the valuable characterizes big data; nevertheless, the overall value of the whole dataset is high, as the volume is immerse, which also supports the need for an appropriate analytical tool.

All this V's are not constant, they varies over time and an industry, they are also interdependent, if the one changes, others will be influenced as well.

It will be a mistake to pay more attention to the first V – volume. Other V's are no less important. As Jagadish (2015, 50) claims, the main reason volume of the data gets more attention is that it is easily measurable, unlike variety and velocity: "I have discussed above, why Volume (or size) gets undue attention. Let me turn now to why I think Variety and Veracity do not get the attention they deserve. One major reason for this lack of attention is that there is no well-accepted measure for either. If there is no measure, it is hard to track progress. If I have a company and develop an innovative system that can handle a slightly larger volume than the competition, I can show this off with measurements against some benchmark. If I am an academic and develop an algorithm that scales better than the competition, I know exactly how to compare my algorithm against the competition and persuade skeptical reviewers. In contrast, consider variety. If I have a product that makes handling variety a little easier, what technical claim can I make that doesn't sound like marketing hype? If I write a paper about a data model that is better at handling variety than the current state of the art, I have to think very hard about how I will compare against the competition and establish the goodness of my idea. Progress is hard in things you cannot measure, in both industry and academia. Variety may be the hardest of the 4Vs to address, but it is the one that people are least motivated to speak about."

There are different techniques for different types of big data being analyzed (structured or unstructured). Types of Big Data analysis methods are as follows: Text analytics, Audio analytics, Video analysis, Social Media analytics, and Predictive analytics.

Audio analytics mostly consist of speech analysis, which is aimed at tracking customer's feedback, as Gandomi (2015, 141) claims it: "Call centers use audio analytics for efficient analysis of thousands or even millions of hours of recorded calls. These techniques help improve customer experience, evaluate agent performance, enhance sales turnover rates, monitor compliance with different policies (e.g., privacy and security policies), gain insight into customer

behavior, and identify product or service issues, among many other tasks. Audio analytics systems can be designed to analyze a live call, formulate cross/up-selling recommendation based on the customer's past and present interactions, and provide feedback to agents in real time."

Video analytics is the least developed brunch, but it bears potential for customer's behavior analysis, as Gandomi (2015, 142) states: "…potential application of video analytics in retail lies in the study of buying behavior of groups. Among family members who shop together, only one interacts with the store at the cash register, causing the traditional systems to miss data on buying patterns of other members. Video analytics can help retailers address this missed opportunity by providing information about the size of the group, the group's demographics, and the individual members' buying behavior."

Text, Social Media, and Predictive analytics are relevant for stock market forecasting, so we will shortly discuss them.

Text analytics deals with all kind of written sources such as news, blogs, emails, documents and so on. Text analytics derives the main ideas out of huge amounts of textual data by creating summaries. Chung (2014) supports the idea that this technique can be used for stock market forecasting, as it can forecast price movements based on financial expert's sentiments.

According to Gandomi A. (2015), Text Analytics techniques include:

1. Information extraction – converting unstructured textual data to constructed one.
2. Text summarization – a technique, which generates meaningful summaries out from texts, using Natural Language Processing methods.
3. Question answering – another technique, using Natural Language Processing Methods. It provides answers to questions, formulated in a natural language, by going through three steps: question processing, text processing, and answer processing.
4. Sentiment analytics – a method, aimed at deriving aggregated customer or expert's opinion regarding some product or events. It operates by classifying opinions as either negative or positive; then, based on the score of these both classes the overall sentiment is determined.

Social media analytics is used primarily for marketing purposes such as customer's satisfaction analysis, community detection, an etc., as social networks provide great opportunities for the target audience analysis. However, it also could be used for stock market forecasting, for example, Antweiler W. (2004) has conducted a study that showed that Yahoo finance message board could be used for stock prices prediction.

Finally, predictive analytics, which includes a variety of quantitative methods that can be used for prediction of almost everything, from the crime rates to the stock market volatility.

Predictive analytics techniques classify in two categories: auto regression and regression analysis. The first type discovers patterns within the chosen variable history; second one exploring dependencies between different variables.

Increased academic attention to the "Big Data" can be explained by the advancement of computation technologies. Modern data mining tools made it possible for researchers to work with huge amounts of structured and unstructured data. Christine E. Earley (2015, 494) supports this statement: "The availability of large amounts of computerized data in companies has been steadily increasing over the years, but recent advances in processing speed, cloud storage, and the rise of social networks has changed the ease of access to data and the nature of data that can be captured and stored for later use. At the same time, software used to analyze large volumes of data (i.e., data mining tools) as well as more sophisticated data visualization tools can potentially increase the ability of individuals to understand the story that the data is telling them".

1.2 Predictive analytics.

Matlis J. (2006, 42) gives the definition of the predictive analytics as follows:" Predictive analytics is the branch of data mining concerned with forecasting probabilities. The technique uses variables that can be measured to predict the future behavior of a person or other entity. Multiple predictors are combined into a predictive model. In predictive modeling, data is collected to create a statistical model, which is tweaked as additional data becomes available."

As it is evident from the definition, predictive analytics uses the same statistical methods as quantitative analysis, but the difference between them is in the sequence of the research steps. Joe F. (2007) describes processes of quantitative analysis and predictive analytics as follows:

Quantitative analysis steps:

1. Theory
2. Hypotheses Development
3. Test

Predictive analytics steps:

1. Data
2. Relationships Development
3. Hypotheses
4. Model Building Test Hypotheses

5. Model Validation

As we can see, Predictive analytics offers more possibilities for analysis, as it can find interdependencies that otherwise could have been overlooked.

The difference between predictive analytics and quantitative analysis can be represented from explanatory vs. predictive modeling perspective.

Explanatory statistical models test predefined hypothesis based on theory. The role of explanatory statistic is to show the casual dependencies between variables. In order to build an explanatory model one should firstly identify the cause and effect relationships between variables, and then build model for testing of his/her hypothesis. In other words, explanatory statistic is used for proving that revealed connections between factors and depending variable are relevant. For an evaluation of such models, analysts use statistical tests, such as, R-squared etc., which measure explanatory power of a model.

Predictive models have different constructing mechanism: instead of focusing on theory based casual links between variables, predictive models are based on association links between variables. Predictive analysis, unlike explanatory, starts with the data. Then it looks for associations between variables within the dataset and build forecasts based on the findings.

Evaluation of predictive models is based on measuring predictive accuracy, instead of explanatory power.

Shmueli (2010) points out four criteria which differs predictive and explanatory analytics: "… causation-association, theory-data, retrospective-prospective, and bias-variance". Bias-variance perspective refers to the different evaluation criteria for predictive and explanatory models: first seeks to minimize sample variance, whereas the latter minimize model's bias.

Both approaches (explanation and prediction) are hardly compatible within a single model, as best explanatory model is not the best predictive one, argues Konishi S. (2007), despite the fact that it has some level of predictive power.

Predictive models increase their accuracy at the cost of higher bias, therefore, prediction models are not necessarily are "true", in a sense that there may not be theoretical foundation for them. Since predictive analytics operates on the big data, it inevitably face challenges, which Fan, J. (2014) has identified as follows:

1. Heterogeneity. Data obtained from the multiple sources and in different formats creates additional difficulties for an analyst.
2. Noise accumulation. Predictive models are build using multiple factors at the same time, and total accumulated mistakes create "noise", which can conceal true influence of some factors.
3. Spurious correlation. Due to huge sizes of the datasets and multiple variables being analyzed, a false correlation may be detected.
4. Incidental endogeneity. It is a threat of breaking one of the traditional assumptions of the regression analysis − exogeneity, meaning that some of the predictive factors could be dependent on the residual term.

Application areas of predictive analytics vary from business related topics, such as retail, marketing and finance, to healthcare and environmental issues. Retailers use predictive analytics to forecast demand on particular product. Marketers use analytics create customers profiles, to determine the reaction of the public on new products, and to detect customer's communities. Law enforcement agencies use it to predict the occurrence of crimes, healthcare systems employs predictive analytics to make diagnosis that is more precise, costume agencies use it for purposes of fraud detection.

There are numerous possible applications for data analytics. Banks and other financial organizations also have much to gain from predictive analytics. Today, big data challenges both firms and individual traders, and those who are capable of rapidly extract relevant information and analyze it, will gain the competitive edge. As the report from SAP (2012) states: "…the profitability keeps falling in recent years, and organizations are now evolving towards smart trading based on big data analytics. Besides designing more complex computing model and system, how to make such large scale computation real time is still a very important problem that is needed to be considered seriously".

1.3 Social Media and Business news Analytics

There is a subset of Big Data, which refers to the big data derived from social media – social big data.

Bello, O. (2016, 47) defines social big data as follows: "Those processes and methods that are designed to provide sensitive and relevant knowledge to any user or company from social media data sources when data sources can be characterized by their different formats and contents, their very large size, and the online or streamed generation of information."

Methods of processing social big data constitute social big data analytics, which is defined by Bello, O. (2016, 47) as follows: "Social big data analytic can be seen as the set of algorithms and methods used to extract relevant knowledge from social media data sources that could provide heterogeneous contents, with very large size, and constantly changing (stream or online data). This is inherently interdisciplinary and spans areas such as data mining, machine learning, statistics, graph mining, information retrieval, and natural language among others. This section provides a description of the basic methods and algorithms related to network analytics, community detection, text analysis, information diffusion, and information fusion, which are the areas currently used to analyze and process information from social-based sources."

Social big data may be of use not only for those companies, who trade in consumer good, but also for financial and for banking sector.

Asset's prices are determined not by impartial machines, but by individuals who trade on the stock exchange. As any human being, they are not completely rational, their decisions are influenced by public's mood and rumors.

Advancement of analytical applications has made it possible for researches to include psychological factors in their predictive model. Tracing these factors is challenging, since they are hidden in the huge amount of unstructured data. One of these factors are customer's sentiments and opinions about a company or a product.

People's expectations and opinions about a particular company or product are reflected in social media platforms, such as Twitter and Facebook. Models for stock prices prediction based on an analysis of public's mood were build and tested in some academic articles, such as Bollen, Mao, and Zeng (2011), and Wu He (2015).

Johan, B. (2011) has shown that even Dow Jones Industry Average index could be predicted by analyzing Twitter mood. First step of constricting predictive model based on information derived from social media is public sentiment's extraction. There are various software tools for that purpose, including IBM Watson. Second step is data processing. It is done by assigning scores or dimensions to every observation. Scores could be "positive", "neutral", "negative", or some other forms. After transforming initial unstructured data into structured scores, usual statistical methods could be applied. Using the same technique, one can build a prediction model based on the machine processing of the vast amount of business news articles. Such a model was build by Chowdhury (2014). Accuracy of forecasting with public's sentiment models is varying from 70 to 80%.

1.4 Market of predictive analytics tools in financial sphere

Market of financial analytics is a fast growing niche. It was estimated in Financial Analytics Market forecast conducted by Research and Markets (2014), that by 2018, total market value of financial analytics will reach the mark of 6.65 billion dollars. Such a rapid growth is driven by the impact of the big data on the operations of banks, audit firms and other types of financial organizations.

Nowadays, researches point out the importance of predictive analytics for all organizations, for example Ventana Research (2016, 3) states: " Organizations increasingly need to understand what's happening right now and to be able to forecast what is likely to happen in both the near future and the long term." As a mean to serve this need, Ventana Research (2016) sees predictive analytics. Currently, there are multiple providers of analytical tools on the market. Among them are such giants as IBM, Microsoft, Google, and Amazon.

According to Doug, H (2015), the leading position on the market of analytics platforms belongs to the IBM Corporation, KNIME, RapidMiner, and SAS. Such a giant as Microsoft is lagging behind, but in a past two years it has showed positive dynamics and now it is catching up with the leaders.

One of the IBM's products became particularly popular among researchers and dada scientists – Statistic Package for Social Sciences (SPSS). SPSS embeds vast arrange of statistical tools and provides customers with the ability to apply econometric modeling to their data. SPSS offers everything the analyst needs, but the main drawback of SPSS is the requirements to the user: only qualified specialist with expertise in statistic and econometric could use SPSS properly.

Apart from SPSS, IBM offers another service, available through cloud – IBM Watson analytics. Watson analytics provides customers with natural predictive and visual analytics. It includes data storage, data processing, data analysis, and visualization. In addition, it can run social media analysis (twitter), helping to assess public's sentiments towards any given event/company/product.

Three key properties of IBM Watson analytics are as follows:

1. Complex arrange of services: unlike other analytical tools, that are supposed to solve particular types of business tasks, Watson analytics helps to refine data, evaluate its quality, analyze it, and create a report, thus rendering use of other tools unnecessary.

2. Predictive analytics: IBM Watson automatically determines the most relevant data, and reveals interconnections between variables.

3. Usage of natural language: IBM Watson allows users to ask questions in common English, thus making it possible for a person without knowledge of statistic science to operate with the data.

Microsoft's, Amazon's, Google's, and SAS's predictive analytics represented by Azure Machine Learning, AWS Machine Learning, Google Predictive API, and SAS Visual Analytics respectively.

In the essence, they are analogs of IBM Watson, all of them provide visualization, analytical, and predictive services, accessible through cloud. Important feature of all these three products is that they offer predefined analytical models for particular business need: banking, insurance, retail etc. Unlike IBM Watson Analytics, they provide customers with the ability to develop their own applications for very specific purposes. There are many other players at the advanced analytics market: Prognoz, Sap, Oracle and so on, but they occupy niche market.

1.5 Research gap.

Influence of the big data and applications of predictive analytics in different spheres of business, healthcare and public safety have gained some attention in the past few years. However, the most attention gained marketing: analyzing customer's feedback, detecting communities via social media, demographical profiling of customers.

Big data and predictive analytics' influence on financial and banking sectors has been noticed in academic circles. There are some academic papers, like Earley, E. (2015), Yoon, H. (2015), and Min, C. (2015) which addressing opportunities and challenges of big data analytics for auditing. Other studies, like Srivastava, U. (2015) analyze the application of big data analytics for banking sector, but they mostly cover customer profiling, risk management, and fraud detection issues. Smith (2015) and Bologa (2010) have discussed the influence of big data and big data analytics on the insurance sector.

Kwan, M. (2014), and Ruta, D. (2014) brought the problem of applicability of big data analytics and predictive analytics for the purposes of increasing effectiveness of trading operations on the stock market to the attention of academics.

However, their research only stated the opportunities and challenges of big data in trading. They did not run empirical check and did not compare analytical platforms, available on the market. Both information deficit and the abundance of information can make it hard for the trader to make a decision regarding his trading strategy. Profit of an individual trader, bank, or broker firm depends on how quickly and effective relevant information is extracted from the high volume datasets of unstructured and structured data. Rise of the big data creates a need to an effective and reliable methods and tools for processing vast amounts of market data.

All of the before mentioned authors have identified possible implications of big data analytics for banking, audit, and insurance, but there is still a place for a research, which goal would be to find out how particular type of financial organization (bank, audit, insurance or trader), could achieve their business objectives using particular types of advanced analytical platforms.

The goal of this research is to fill the research gap by assessing possible applications of predictive analytics for stock market forecasting.

1.6 Research methodology and organization of the study.

In a course of this research, we will use quantitative methods to analyze and compare forecasting abilities of the leaders of the market of advanced analytical platforms: IBM Watson Analytics, SAS Analytics, KNIME, and RapidMiner. The comparative analysis will be based on the set of predefined KPI's.

Using the KPI's, we will assess the ability of these analytical platforms to execute business tasks of financial organizations. Based on the analysis, we will run a comparative analysis of the platforms and generate recommendations regarding which platform to use for the purposes of stock market forecasting.

Special attention will be paid to one of the KPIs – forecasting accuracy. For the purposes of comparing chosen analytical platforms by this KPI, we will build predictive models in IBM Watson Analytics and Gretl statistical package.

Firstly, we will build econometrical models for prediction of currencies exchange rates. For that purpose, we will build two types of models: ARIMA and factor regression, which use prices of the main export product of the country.

Next financial asset's price we will try to predict is blue chips of stock markets – IBM, Microsoft, P&G etc. As a theoretical base, we will use Capital Asset Pricing Model – CAPM. United States financial market is one of the most developed ones, therefore it's reality is as close to the Effective Market Hypothesis (EMH) as it gets on real life markets.

The last financial assets we will take into consideration are stock indexes. The importance of considering stock indexes is driven by the fact that they serve as a guideline for traders, analyst and investors, because they reflect overall situation on the market.

It is the first phase of the empirical research, and it will be conducted using Gretl statistical package. Our next step will be the construction of predictive models for the same assets using the same datasets in all aforementioned analytical platforms. Apart from building alternative quantitative models using financial data, we will make use of social big data, by running twitter analysis with the help of IBM Watson analytics.

Accuracy of forecasts will be assessed through two characteristics: Mean Absolute Percentage Errors and the potential profitability of applying such models. Potential profitability will be estimated during the simulation experiments. We will imitate real life trading using given models. We will set an investor's behavior as follows: investor is profiting from the difference between prices of the same assets in two consequent time periods. If the model predicts that the price will go down, the investor buy the asset, with an intention to sell in the next period regardless of its actual price. If the model predicts depreciation of the asset, than it goes vice versa.

The result of the research will be a comparative analysis of forecasting abilities of some the main analytical platforms available on the market.

1.7 Conclusion of Chapter I.

The rise of the big data in the recent years has created challenges and opportunities for every type of business. In order to tackle this challenges and not to miss the opportunities it's necessary to use predictive analytics techniques. Big data consist of vast amount of structured and unstructured information. It is characterized by the three V's – volume, variety, and velocity. For different kinds of big data being analyzed there is different type of data analytics techniques. Data analytics consists of text analytics, audio analytics, social media analytics, and predictive analytics.

Big data affects many spheres of business, including trading, as it allows processing and analyzing of immense amounts of data, thus making it possible for analyst to uncover interdependencies and patterns, which otherwise would have been ignored. Big data holds potential to increase effectiveness of trading deals on the stock market, therefore it is subject of interest for both individual traders and broker firms. Trader's interest in the analytical platform is its capabilities to explore the data, to find out interrelationships and correlations between variables.

These interdependencies and correlations within a dataset could be detected using traditional statistical methods. However, predictive analytics and conventional statistical methods are not completely similar, despite the fact, that predictive analytics and econometrics use the same mathematical and statistical toolkit. There is one fundamental difference between them: in order to build econometrical models, one should find theoretical grounds for it, formulate statistically verifiable hypothesis, and test it. This approach leads to a creation of an explanatory models, which describe factors that drives observable variable, however, this kind of models don't have the best predictive accuracy. Predictive analytics, just like econometric modeling, uses statistic methods, but it differs in the research approach. Predictive analytics doesn't need to test predefined hypothesis, instead of doing so, it explores interdependencies between observable variable and whole set of possible predictive factors. As a result, a predictive model is created, which however, may lack theoretical explanation and which could be more biased than explanatory one. Additionally, advancement of cloud computing made it possible to run social media and investor's sentiment analysis.

Finally, such characteristics of an analytical platform as text analysis and social media analysis is an object of interest for every financial organization (except for audit firms, since the applicability of social media to the audit isn't confirmed), as the majority of information comes in an unstructured form.

There are many analytical platforms available on the market, we will take into account only top five of them, according to Gartner's Magic Quadrant of advanced analytical platforms (2016). Most of them are available only through cloud (IBM Watson, Azure Machine Learning, SAS Visual analytics, Amazon Machine learning), however some platforms offers their services offline: KNIME, RapidMiner, which were recognized by Gartner as one of the market leaders.

Market of advanced analytical platforms is one of the most dynamic. Comparison of Gartner's Magic Quadrants from 2014 and 2015 reveals serious movements on the market. However, one player on the market attracts special attention – IBM, with its cloud-based analytical service called Watson Analytics. IBM has been the leader of the market for several years, and its service provides an easy way for a researcher to analyze and visualize huge amounts of data.

Ability to simultaneously process big datasets holds the potential for stock market analysis. Nowadays, there are too many information on the market, coming from multiple sources, its impossible to assess all relevant information in a short time, and time is of the essence when it comes to forecasting a stock market.

Chapter II. Research framework.

2.1 Research goals, KPIs, objectives, questions, and limitations.

Purpose of this work is to provide potentially interested parties (trading firms), with the comparative analysis of predictive analytics tools and providers, in order to help them to make a decision regarding which product to use for a particular task.

In a course of this research, we will analyze and compare main advanced analytical platforms that are available on the market. Each advanced analytical platform has its own characteristics that are identified by the Ventana Research (2016) as follows:

1. User roles and self-service: this characteristic reflects the ability of a platform to be used by different kind of users, with different data analysis capabilities and different requirements to analytics.

2. Information Optimization: it reflects the ability of an analytical platform to manage different kinds of data flows that are coming from different sources, and the ability to refine the data.

3. Range of analytical capabilities: it includes visualization capabilities, data exploration capabilities (uncovering of hidden patterns), and ability to detect particular events in the dataset.

4. Cloud and Mobile deployment.

5. Time to Value: the ability of a platform to perform the analysis and present the result in the shortest time possible.

From these five KPI's we will take two: User roles, and self-services, and Range of analytical Capabilities. We will break down them into sub criteria as follows: Visualization, Simplicity of Use, Predictive Analytics capabilities, Range of Econometric Modeling, Textual Analytics capabilities, and Social Analytics Capabilities.

After evaluating analytical platforms using these KPI, we will analyze how well each of them addresses the needs of particular kind of financial organization. Then we will provide interested parties with the recommendations regarding which advanced analytical platform to use for each of the business objectives.

In addition, we will look into how the ability of an advanced analytical platform (using IBM Watson Analytics) to suggest predictive models compares with standard theoretical approaches to stock price forecasting.

Goal of this research is twofold. Firstly, it is to run a comparative analysis of main advanced analytical platforms. Secondly, it is to assess the ability of IBM Watson Analytics to suggest effective predictive models for stock price forecasting.

Research questions of this work are as follows:

1. Which analytical platforms is a better fit for the purposes of stock price forecasting?
2. Does analytical platform (Using IBM Watson Analytics as example) suggest effective predictive models for stock forecasting, in comparison with standard theoretically based econometric models?

Research objectives:

1. To evaluate analytical platforms (IBW Watson Analytics, SAS Analytics, KNIME, and RapidMiner), using KPIs mentioned before.
2. To make a comparative analysis of the analytical platforms.
3. To rank them based on their ability to make predictive models for stock market forecasting.
4. To construct and evaluate theoretically based econometric models for stock prices forecasting.
5. To construct econometric models for stock price forecasting using factors, suggested by IBM Watson Analytics Prediction function.
6. To compare the performance of theoretically based, and Watson Analytics suggested models.

Analytical platforms that will be taken into consideration: (IBW Watson Analytics, SAS Analytics, KNIME, and RapidMiner). This choice is justified by the Gartner's Magic Quadrant of advanced analytical platforms (2016), which has identified them as the market leaders.

Limitations:

1. Not all advanced analytical platforms available on the market are considered.

2. Ability of IBM Watson Analytics to suggest predictive models will be compared only with mostly common used econometric models: comparative analysis with all econometric possible econometric models is impossible, as there are too many of them, and new ones could always be generated.

3. Not all analytical capabilities of analytical platforms will be empirically tested.

4. Simulation of the potential profitability is made under the assumption that an investor have access to all necessary information and reacts on it instantly.

2.2 Methods of evaluation of advanced analytical platforms.

Evaluation of the analytical platforms will be done using Analytical Hierarchy process. According to Abdullah (2013), AHP is conducted through seven steps:

1. Determination of the hierarchy of criteria and calculation of the normalized matrix.

2. Determination of criteria weights

3. Determination of the eigenvector.

4. Check of the consistency ration.

5. Comparison of the alternatives.

6. Calculation of the alternative's scores

7. Ranking of alternatives.

Hierarchy of criteria is determined by the relative importance of them for the goal (car purchasing, vendor choice etc.). In a result of a pairwise comparison of the criteria, a matrix n x n is created. Its elements reflect the relative value of different criteria to each other. For example, element a_{ij} indicates value of "i" criteria to "j" criteria. $a_{ii}=1,$ and $a_{ji}=1/\ a_{ij}$.

Next, a normalized matrix is defined. Element a_{ii} in this matrix is determined as the results of dividing the values, derived in the result of pairwise comparisons of row "i" relative to column "i", by the sum of the pairwise comparisons in the "i" column.

The criteria weight is determined as a mean of elements of normalized matrix:

$$\sum_{i=1}^{n} \mu_i = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} (1)$$

Eigenvector is determined as follows:

$$w_i = \frac{\sqrt[n]{\mu_i}}{\sum\limits_{i=1}^{n} \mu_i}(2)$$

Consistency ratio is calculated as:

$$CR = \frac{CI}{RI}(3)$$

$$CI = \frac{\gamma_{max} - n}{n-1}(4)$$

$$\gamma_{max} = \sum_{i}^{n} \frac{A w_i}{n w_i}(5)$$

RI is a random index, which takes values depending on the number of elements (n). CR should me no more than 0.1.

Scores of the alternatives are calculated using this equation:

$$Ascore = max \sum_{j=1}^{n} a_{ij} w_i (6)$$

Based on this scores, final ranking is constructed.

2.3 Methods of comparing the forecasting accuracy of IBM Watson and statistical packages.

Method of research: comparative analysis based on the results of generated by quantitative methods. For the purpose of comparing forecast accuracy of different tools, we will consequently create predictive models in statistical package and IBM Watson using the same dataset.

In order to have the most tried and reliable econometric models for comparison, we will run forecasts of currencies exchange rates, forecasts of stock's price dynamic, and forecasts of stock indexes.

Econometric model building follows three steps:

1. Theory
2. Hypotheses Development
3. Test

We will forecast currencies exchange rates using two approaches: ARIMA models, and linear regression models, which uses prices of the most exported commodities as an independent variable. Theoretical foundation of these models could be found in Meese, R., Rogoff (1983). And Rogoff, Rossi (2015).

Additionally, we will build CAPM models for stocks of the biggest corporations, such as Google, Microsoft etc. We will use only USA stock market for building CAPM models, because

CAPM operates under the assumptions of EMH (Effective Market Hypothesis), therefore, CAPM doesn't fit for developing stock markets.

Forecasting with the IBM Watson Analytics differs from building econometrical models, the main difference is that it doesn't require strong theoretical grounds in order to make a model – it analyzes the whole dataset and automatically suggests.

Predictive analytics steps:

1. Data
2. Relationships Development
3. Hypotheses
4. Model Building Test Hypotheses

5. Model Validation

As we can see, IBM Watson lacks theoretical grounds for model building, but best predictive models are not necessarily the best theoretically based as it is stated by Shmueli, G (2010).

Comparison of the predictive models will be based on the two indicators:

1. Mean Absolute Percentage errors

2. Potential profitability

Potential profitability will be estimated as profit, generated by the given model during the simulation.

Simulation will be run in accordance with rules as follows:

1. If model predicts, that price of the asset will rise in the next period, an investor makes a decision to buy the asset.

2. If model predicts, that price of the asset will fall in the next period, an investor makes a decision to sell the asset.

3. If an investor bought the asset, he would sell it in the next period regardless of of its new price.

4. If an investor sold the asset, he would buy it back in the next period, regardless of its new price.

At the end of the prechosen period, investors stops and calculates his/her returns, which will be used as an indicator of forecasting accuracy of the model. In order to have more reliable indicator of the forecasting accuracy, we will run a model, simulating real life trading.

Rules of the model are simple, if it anticipates, that asset's price will increase in the next period, than an investor takes the decision to buy the asset, with the intention to sell it afterwards. Depending of the actual change of the prices, such operations could bring profits or loses.

2.4 Method of currency exchange rate forecasting using Statistical Packages

Since the publication of a highly cited article of Meese, R. (1983), it has become a sort of benchmark to compare all currency exchange rates models with the Random Walk models, which performs no worse than any other model.

However, some more recent researches like Moosa, I. (2014), argue that unbeatable random walk is, in fact, an illusion. They argue, that random walk model seems superior only if it is evaluated it in terms of mean square error, absolute square error and rood mean square error, but if model is evaluated by its direction forecasting power and profitability, than random walk lose it to almost all other models.

Random Walk is a type non-stationary time series, which is defined as follows:

$$X_t = X_{t-1} + e_t \quad (7)$$

Where X_t is an observable variable, and e_t is a pure random component.

The difference between the random walk and auto regression AR (1) is that an effect of every random component is preserved forever.

If the process begins with t=0, than:

$$X_t = X_0 + e_1 + \ldots e_t \quad (8)$$

In a more general case, there is a constant B_1, which turns the process into a random walk with a trend:

$$X_t = X_0 + B_1 t + e_1 + \ldots e_t \quad (9)$$

Another nonstationary process is a time series with the determined trend:

$$X_t = B_1 + B_2 t + e_1 \quad (10)$$

The main difference between this model and the random walk is that time series with determined trend has a tendency to return to the trend's line, while random walk with trend doesn't necessarily returns to the trend's line.

One more approach to currency exchange rate forecasting is a regression model, based on the prices of main export commodities of the given country. Such a model was tested by Ferraro, Dominico, F. (2015). They tried to forecast US. dollar - Canadian dollar, US. dollar –Australian dollar, US. dollar –Norwegian krone, US. dollar – South African rand, and US. dollar – Chilean peso, currency exchange rates based on the prices of oil, gold, and copper. The results revealed short-term relationships between prices of country's main commodity price and its currency nominal exchange rate.

However, the applicability of such a model is limited by countries, which have small number of main export commodities, meaning that most of developed countries currencies exchange rates couldn't be predicted using this model.

As an approach to currency exchange rate forecasting, we will use ARMA and ARIMA auto regression models.

ARMA model is defined as follows:

$$X_t = B_1 + B_2 X_{t-1} + \ldots + B_{p+1} X_{t-p} + e_t + a_2 e_{t-1} + \ldots + a_{q+1} e_{t-q} \quad (11)$$

Where X_t – observable variable, B_t – coefficient, which determines the influence of the previous observations on the current one, $a_t - ¿$ a coefficient, which determines

2.5 Methods of stock forecasting using Statistical packages

Econometrical methods of financial market analysis have strong mathematical and statistical grounding. However, their applicability is limited due to assumptions, upon which econometrical models are based.

Most of theoretical models of stock market forecasting require so-called Efficient Market Hypothesis.

Efficient market hypothesis refers mostly to the information effectiveness of a market. Efficient market hypothesis implies that information is equally available to all participants of the market; they interpret it in a similar manner and instantly use it to adjust their strategy and operations.

In addition, efficient market theory suggest that all players are rational, have similar goals and use similar strategies.

Main characteristic of an efficient market is a result of the realization of all aforementioned assumptions. If a market is efficient, then prices of assets instantly, completely and correctly assimilate all available and relevant information, and reach equilibrium, thus making regular gain of abnormal incomes impossible.

In the efficient market, it is considered that expected returns includes all systematic risks, and provide investors with acceptable returns, consistent with all other similar risk level assets.

One of basis models, based on efficient market theory is Capital Assets Pricing Model (CAPM). Its main equation looks as follows:

$$\mu_i = R_0 + \beta_i \left(\mu_M - R_0 \right) \quad (11)$$

Where $\mu_i - ¿$ expected return of any given asset; R_0 – risk free return, β_i – beta coefficient, reflecting the nature of the asset (riskier and more profitable assets have $\beta_i > 1$; and less risky and less profitable ones have $\beta_i < 1$); μ_M – average market return on assets.

CAPM is based on the list of assumptions:

1. Investors evaluate assets using their expected returns and risks
2. Expected returns are stochastic
3. Risk is measured as dispersion of returns

4. Investors are trying to maximize their asset's returns

5. Investors are risk aversive

6. Absence of a monopolistic influence on the market

7. Absence of taxes

8. Absence of transactional costs

9. Absence of unexpected inflation

10. Assets are infinitely divisible

11. No limitations on leasing and lending on risk free rate

12. All investors have similar planning horizon

13. All investors evaluate probability distribution of expected returns

14. Information is free and all investors have equal access to it.

It is evident from the list that CAPM assumptions are unrealistic, as nearly the half of them contradict the reality of actual financial market. However, this model serve as a base from which other, more realistic models could be derived. It is done by loosening some of the aforementioned assumptions, thus making model more applicable for actual forecasting.

Another class of econometric models is factor models. These models assume that expected return of an asset could be determined as a reaction to a change of some economic factors, such as GDP, inflation or oil prices.

Factor model tries to consider main economic factors, influencing prices of assets. It implies that any two given stocks are correlated with each other only through common economic factors. Every factor, influencing expected return of a given asset, which is not in the model, considered unique; therefore it doesn't correlate with unique factors of other assets.

2.6   Conclusion of Chapter 2.

In a course of this research, we will evaluate the abilities of top analytical platforms (IBM Watson Analytics, SAS Analytics, KNIME, and RapidMiner), to serve the needs of banks, audit firms, insurance companies, and trades by assessing these platforms using the set of two main KPIs: User-Friendliness, and Range of Analytical capabilities. These main KPI's are subdivided into six criteria: Visualization, Simplicity of Use, Predictive Analytics, Econometric Modeling, Textual Analytics, and Social Media Analytics.

Then we will evaluate chosen analytical platforms using Analytical Hierarchy process and the set of KPI's mentioned above. AHP will performed through a series of pairwise comparisons, which will determine the relatives weights of criteria and ranking of the alternatives (IBM Watson Analytics, SAS Analytics, KNIME, and RapidMiner). After

conducting AHP, we will determine the most appropriate analytical platform for the purposes of stock market forecasting.

In case of statistical packages, we will used theoretically based econometric models, and in case of IBM Watson analytics we will let the platform to suggest optimal models by itself. This approach a potential problem: lack of the theoretical groundings. For a trader, it may appear be irrelevant, since he/she mostly cares about the accuracy of forecasts, however, without theoretical basis it is impossible to guarantee the stability of the model: it could have just happened that the factors, which affected the predicted variables, are spuriously correlated.

In the research, we will build series of models. The first one will be standard random walk models for currency's exchange rates. It will used for a comparison with other models, since they will make any sense only in case if they outperform the random wall.

Another series of predictive models for currency's exchange rates will be constructed using simple one-factor model that use price of the most exported commodity as a predictor. The dynamic of stock market will be analyzed by applying Capital Asset Pricing Model to the blue chips of the United States stock exchange: Microsoft, Apple, IBM, Bank of America, Walmart, and P&G. The US stock market was chosen because of the necessity of operating under the Effective Market Hypothesis, which more likely to be true in the developed market, rather than the emerging one.

Final series of predictive models will be constructed in Gretl, but in this case, factors will be chosen based on the suggestions of IBM Watson Analytics, which automatically determines drivers of a given variable.

Predictive accuracy of the forecasts generated by aforementioned models will be estimated by two characteristics: Mean Absolute Percentage Errors, and potential profitability. The latter characteristic will be assessed through the results of a trading simulation experiment during which we will imitate real-life trading using all of the models we have constructed.

Chapter 3. Empirical estimation of analytical platforms.

3.1 Evaluation of the Analytical Platforms

3.1.1 Justification of the choice of analytical platforms taken for consideration

According to the Gartner Magic Quadrant for advanced analytical platforms (2016), the market is divided into four categories: Leader, Challengers, Visionaries, and Niche players. This classification is based on their abilities to execute (performance metrics) and their completeness of vision, which could be interpreted as future perspectives. See figure 3.

There are several vendors and analytical products in each category. In this research we focus our attention on the leaders and visionaries. Among the market leaders are IBM, SAS, Dell, KNIME, and RapidMiner. Visionaries are represented by Microsoft, Alteryx, Alpine Data, and Predixion Software.

Another research – Forrester Wave (2015), suggest a different picture, based on the current performance and strategy: IBM and SAS remain the leaders. However, KNIME, RapidMiner and Dell are removed from the leaders section and ranked as strong performers. See figure 4.

According to both reports, IBM is the market leader: it shares its place with SAS, but according to Forrester Wave (2016), it has better perspective for future (higher strategy rank). KNIME and RapidMiner occupy similar positions in both rankings, also they offer similar approaches to analytics – both are available offline and provide clients with cost-benefit ratio, as it states Piatetsky (2016). So, we will chose for platforms for further analysis: IBM Wastson, SAS, KNIME, and RapidMiner. Dell is set aside, because it is noticeably behind other leaders.

Figure 3. Gartner's Magic Quadrant for Advanced Analytical platforms 2016
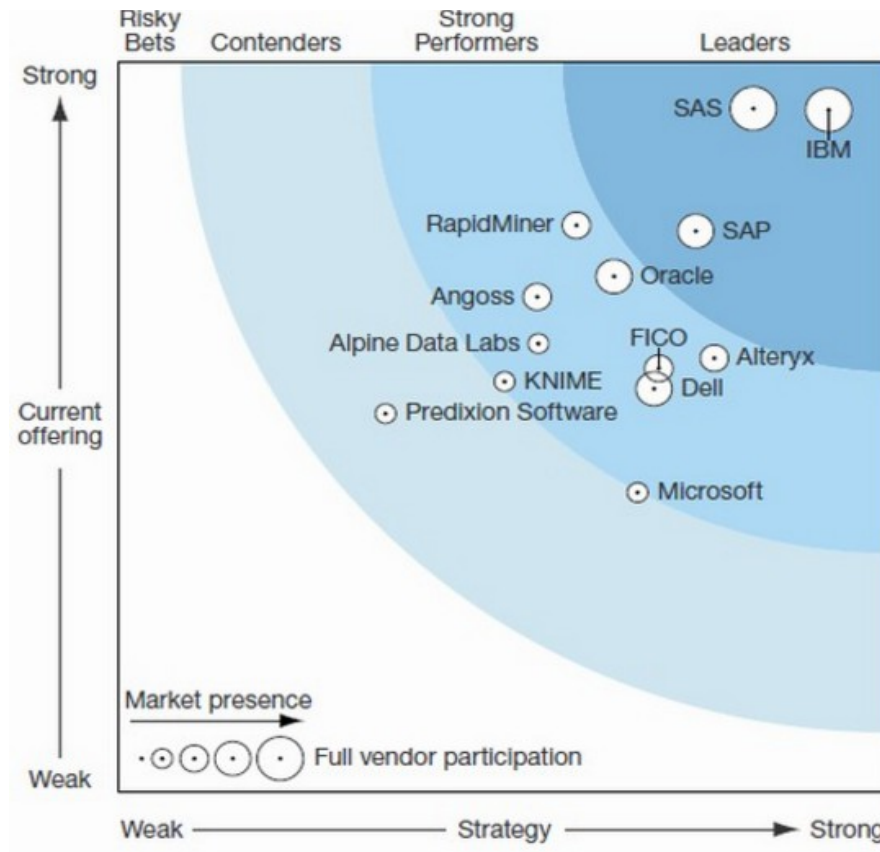
Figure 4. Forrester Wave 2015

3.1.2 Results of Evaluation of Analytical Platforms

We will apply simple Analytical Hierarchy Process Method, using BPMSG AHP Online System (http://bpmsg.com/academic/ahp.php). Goal of AHP is to choose most appropriate analytical platform for stock price forecasting. According to Gartner's Magic Quadrant for Advanced Analytics (09 February 2016, ID: G00275788) the key alternatives are IBM Watson, SAS, KNIME, and RapidMiner. We have two main criteria: User-friendliness and Range of Analytical Capabilities. These criteria could be broken down into sub-criteria as follows:

1. User-Friendliness: Visualization, and Simplicity of Use.
2. Range of Analytical Capabilities: Predictive Analytics, Econometric Modeling, Textual Analytics, and Social Media Analytics.

The relative weights of these criteria were determined through a series of pairwise comparisons with each other. The comparisons were made as follows:

1. Range of Analytical Capabilities is more important than User-Friendliness.
2. Simplicity of use is more important than Visualization.
3. Predictive Analytics is equally important to Econometric Modeling; Predictive Analytics is more important than Textual Analytics and Social Media Analytics; Econometric Modeling is more important than Textual Analytics and Social

Media Analytics; Textual Analytics is equally important to Social Media Analytics.

Description of these criteria is presented in the Table 1. And results of this pairwise comparison in BPMSG AHP Online System are presented in the Table 2.

Table 1. Description of the criteria.

Criteria	Description
Visualization	Refers to quality of data and analysis visualization, provided by a platform. requirements of IT and statistical expertise.
Simplicity of Use	Refers to the requirements of IT and statistical expertise.
Predictive Analytics	The ability to suggest predictive factors.
Econometric Modeling	Range of the statistical and econometrical tools, which a platform provides with.
Textual Analytics	Reflects the range of textual analytics techniques, provided by a platform.
Social Media Analytics	Reflects mostly the range of social media and news sources, which a platform is capable of analyzing

Table 2. Decision Hierarchy

Level 0	Level 1	Level 2	Global Priorities
Analytical Platform	User-Friendliness	Visualization	11.1 %
		Simplicity if Use	22.2 %
	Range of Analytical Capabilities	Predictive Analytics	22.2 %
		Econometric Modeling	22.2 %
		Textual Analytics	11.1 %
		Social Media Analytics	11.1 %

Our next step is to evaluate the alternatives using these criteria. The results of the evaluation are shown in the Table 3.

Table 3. Evaluation of the Platforms (sources: Bloor Group, KMIME Documentation, SAS Product Documentation, and RapidMiner Documentation).

Visualization			Comments
Platform	Priority	Rank	
SAS	39.5%	1	This ranking is based on the how well Visualization
IBM Watson	27.8%	2	is integrated into the analytical process, and what

Platform	Priority	Rank	
KNIME	16.3%	3	
RapidMiner	16.3%	3	
Simplicity			
Platform	Priority	Rank	
IBM Watson	39.5%	1	IBM occupies the first place, because it doesn't require deep statistical expertise from the user, and offers simple interface. The reason why both KNIME and RapidMiner hold 3rd rank, is that they require some level of statistic expertise, and have more complicated interface.
SAS	27.8%	2	
KNIME	16.3%	3	
RapidMiner	16.3%	3	
Predictive Analytics			
Platform	Priority	Rank	
IBM Watson	30.0%	1	Both IBM Watson and KNIME directly states the predictive analytics function.
SAS	30.0%	1	
KNIME	20.0%	2	
RapidMiner	20.0%	2	
Econometric Modeling			
Platform	Priority	Rank	
SAS	28.6%	1	All platforms except for IBM Watson offers broad range of econometrical and statistical models, while IBM Watson has replaced It with Data Exploration and Predictive functions.
KNIME	28.6%	1	
RapidMiner	28.6%	1	
IBM Watson	14.3%	2	
Textual Analytics			
Platform	Priority	Rank	
IBM Watson	40.0%	1	All of the platforms offers textual analytics functions, but IBM Watson is the only one capable of answering questions, formulated in natural language.
SAS	20.0%	2	
KNIME	20.0%	2	
RapidMiner	20.0%	2	
Social Media Analytics			
Platform	Priority	Rank	
IBM Watson	28.6%	1	All platforms have social analytics functions; however, RapidMiner can analyze only twitter.
SAS	28.6%	1	
KNIME	28.6%	1	
RapidMiner	14.3%	2	

After evaluating the alternatives (IBM Watson, SAS, KNIME, and RapidMiner) in the BPMSG AHP Online System, we have the results, which are presented in the Table 4.

Table 4. Ranking of Analytical platforms.

	Priority	Rank
IBM Watson	29.4%	1
SAS	29.0 %	2
KNIME	21.6%	3
RapidMiner	20.0%	4

As we can see it in the Table 3. IBM and SAS are almost similar in the regard of suitability for stock price forecasting, according to the AHP method. Overall, the result is consistent with Gartner's Magic Quadrant and Forrester Wave. However, IBM Watson has

scored a bit better, that is why we will use it for our further analysis, which is presented in the next chapter.

3.2 Evaluation of the forecasting accuracy of IBM Watson Analytics

3.2.1 Data description

In the Table 5, we can see a description of the data we will use in the stock market forecasting experiments. Variables are classified into four categories: stock prices, prices of resources (gold, oil, and natural gas), values of the market indexes, and currency's exchange rates. Observations cover the period from 01.30.2015 to 01.04.2016

We will use two types of software to run the predictive modeling: Gretl statistical package and IBM Watson Analytics. Type of models, which is marked as IBM+Gretl in the Table 7, was build in a steps as follows: after uploading the dataset to IBM Watson Analytics, the predictive function was applied. It has suggested predicting factors for each target variable (stocks and currency's exchange rates), after that, simple two-factor regression models were build in the SPSS, using suggested by the IBM Watson Analytics predictive factors as independent variables. The random walk models are basically just ARIMA (0,1,0) models. They will be used just as a basis for comparison.

Table 5. Data description (Source: Finam)XSoftware	Model	Variables	Number of observations
Gretl	Random Walk Models	ERO/USD	429
		USD/CAD	426
		USD/YEN	428
		USD/ZAR	426
		USD/NOR	425
		USD/CNY	393
		USD/RUB	426
	One-Factor models	USD/NOR	363
		USD/ZAR	
		USD/RUB	
		BRENT	
		Gold	
	CAPM	S&P 500	286
		BAC	
		IBM	
		MSFT	
		P&G	
		Walmart	
		Apple	

IBM+Gretl	Two-Factor models	S&P 500	225
		DJI	
		RTS	
		Nikkei	
		CSI	
		FTSE	
		Shanghai	
		NASDAQ	
		Gold	
		Natural Gas	
		Brent	
		ERO/USD	
		USD/CAD	
		USD/YEN	
		USD/ZAR	
		USD/NOR	
		USD/CNY	
		USD/RUB	
		Exxon Mobil	
		Chevron	
		BAC	
		IBM	
		P&G	
		Walmart	
		Apple	

One-factor models are predicting the currency's exchange rates based on the prices of the most exported commodities (oil, and gold). CAPMs predict the prices of the stock. It was build using week prices of the "blue chips" of the US stock market. Role of the average market indicator was played by the S&P 500 index. Interest rate of the 4 week reassure bills was used as the risk free rate (Rfr=2%). Return on assets is calculated as the difference between stock's price in moment t and stock's price in moment t-1, divided by the stock price in the moment t-1:

$$R = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (12)$$

Specifications of the models are shown in the Appendix 1 and Appendix 2.

3.2.2 Forecasting stock prices with theoretically based models.

3.2.2.1 Results of the Random walk models for currencies.

Random walk model is the basis for comparison for any other forecasting model, as any predictive model makes sense only if it beats the random walk. Using Gretl statistical and econometrical package, we have built ARIMA(0,1,0) time series models, which are equivalent to

the simple Random walk. In the Table 6, we can see the error metrics for the random walk model for currencies exchange rates.

Table 6. Percentage errors of Random Walk models

Model	MPE	MAPE
EUR/USD	0.004209	0.004209
USD/CAD	-0.0011137	0.056509
USD/NOK	-0.077933	0,077933
USD/RUB	-0.27649	0,27649
USD/ZAR	8.1111	8.1111
USD/CNY	-0.0089793	0.0089793
USD/JPY	-0.032764	0.032764

As we can see in the Table 2, random walk models have produced quite small mean percentage errors and mean absolute percentage errors, with the exception for USD to South African Rand Exchange rate (ZAR). It might give an impression that random walk performs greatly, however, as it is supported by Elliot, G (2013), for the purposes of profiting from the differences in the exchange rates, it's more important to foresee the direction of change, rather than to give more accurate estimation. Low percentage error in Random Walk case could be caused by the fact that the forecasted value differs from the previous observation only by small random value.

Our next step is to estimate potential profitability of trading main currencies using random walk model. For that purpose, we have run the simulation test in Excel 2013, using rules as follows: if the investor expects appreciation of the asset, then he buys it, and vice versa. The results are shown in the Table 7. We have used 30 last forecasted values of each currency's exchange rates, for an imitation of real life trading.

Table 7. Results of the simulation of Random Walk

Trading simulation (Random Walk)	
Model	Profitability
EUR/USD	0,39%
USD/CAD	-6,20%
USD/NOK	-1,38%
USD/RUB	-17,45%
USD/ZAR	-0,46%
USD/CNY	0,96%
USD/JPY	1,64%

As expected, results of the simulation reveal that Random walk model is absolutely unfit for trading, in 4 out of 7 cases, the profitability is negative, especially in case of Ruble, which

has shown over -17% loses. Even positive examples have very low profitability. The average return is -3.2%, and if it were real life trading, than the loses would be even bigger, as there are transaction costs and time lags. Thus, it is safe to conclude that random walk model is completely unfit for real life application.

3.2.2.2 Currency's exchange rates forecasting using factor models.

Table 8 presents the description of factor models. In accordance with Dominico, F. (2015) we have built predictive model for currency's exchange rate forecasting using prices of mostly traded commodities as predictors. Models were built in Gretl econometrical package using "ordinary least square" option.

Table 8. Description of factor models for currencies.

Model	Model's Parameters		Model's Statistic		
	Coefficient	Sig,	R-squared	MPE	MAPE
USD/NOK			0,79542	-3,6611	3,6611
const	10,1061	<0,0001			
Brent	−0,039167	<0,0001			
USD/RUB			0,770869	4,0127	4,0127
const	83,0842	<0,0001			
Brent	−0,48308	<0,0001			
USD/ZAR			0,169913	14,503	14,503
const	26,4943	<0,0001			
Gold	−0,011423	0,0003			
USD/CAD			0,771673	-0,24727	4,19840
const	1,4658	<0,0001			
Brent	−0,00414812	<0,0001			

All factors are statistically significant and they have expected influence on every currency (the higher the price of the commodity, the lower USD exchange rate). However, these models demonstrate bigger mean percentage errors than the random walk. In that sense, they don't beat the random walk.

Three out of four models have high R-squared (>0.7), which implies good explanatory power of models. The only exception is USD/ZAR model, which has very low R-squared (=0.169) and the highest Mean Absolute Percentage Error (14%). This result leads us to the thoughts that, gold isn't the main export product in South Africa anymore.

Our next step is to estimate potential profitability of trading main currencies using simple one factor regression. For that purpose, we have run the simulation test in Excel 2013. We have used 30 last forecasted values of each currency's exchange rates, for an imitation of real life trading.

As it is shown in the Table 9, trading with factor models brings way higher returns, that just random walk, because factor models manage to generate more accurate predictions of the direction of price's change. Average return for this model is 26%.

Table 9. Results of the simulation of the factor models.

Trading simulation (Factor regression)		
Model	Factors	Profitability
USD/CAD	Brent	0,2151143
USD/RUB	Brent	0,3132015
USD/ZAR	Gold	0,2284068
USD/NOK	Brent	0,2875645

3.2.2.3 Stock forecasting using CAPM model.

Using "ordinary least square" function in Gretl statistical package, we have built CAPM for every of stocks as follows: Apple, IBM, Microsoft, Procter & Gamble, Walmart, and Bank of America. As a factor we have used the risk premium:

$$RP = \left(\mu_M - R_0 \right) \quad (13)$$

Where μ_M is return on S&P index, and R_0 is four weeks treasure bill interest rate.

As we can see it in the Table 10, CAPM model produce quite poor results both in terms of explanatory power (low R-squared) and accuracy of forecasts, sometimes mean percentage errors exceed 100% (Average MAPE = 177%), meaning that the forecasts is radically different with the reality. Despite the fact that in all cases, risk premium as a factor was significant, and R-squared is tolerable (except for Walmart case), the models appear to be unfit for the actual forecasting. Because of huge deviations of forecasted values from the actual ones.

Table 10. Description of CAPM for stocks.

Model	Model Parameters		Model's Statistic		
	Coefficient	Sig.	R-squared	MPE	MAPE
Bank Of America			0.549330	574.13	574.13
const	0.000181202	<0.0001			
RP	1.32574	<0.0001			
Microsoft			0.464251	202.74	258.91
const	0.00130152	0.0833			
RP	1.13318	<0.0001			
Apple					
const	−0.00068693	<0.0001			
SP	1.14488	<0.0001			
Walmart			0.214836	59.562	59.562
const	−0.0015747	0.0246			

RP	0.592545	<0.0001			
IBM			0.453587	72.152	72.152
const	−0.000336033	<0.0001			
RP	0.919689	<0.0001			
P&G			0.459370	66.181	66.181
const	−0.000790775	0.0787			
RP	0.671332	<0.0001			

Our next step is to estimate potential profitability of trading blue chips stocks using CAPM model. For that purpose, we have run the simulation test in Excel 2013. We have used 30 last forecasted values of each currency's exchange rates, for an imitation of real life trading.

Results of the trading simulation (Table 13) confirms the point that CAPM is unfit for stock market forecasting. CAPM has generated significant potential outcome only in 2 out of 6 cases, in two cases, the results were negative, and the last two have demonstrated negligible profits, which would not even cover transactional costs. Average return is 5%, which demonstrates that despite huge deviations of forecasted values from actual ones, in some cases CAPM still correctly predict the direction of change.

Table 11. Results of the CAPM simulation.

Model	Profitability
BAC	0,1774171
IBM	0,19201815
MSFT	0,03278492
P&G	-0,0342963
Walmart	0,01295297
Apple	-0,0838422

3.2.3 Forecasting stock market using IBM Watson analytics.

3.2.3.1 Models for stock forecasting.

We have used free version of IBM Watson analytics to conduct our experiment. After uploading our dataset consisting of 26 variables, IBM Watson Predict option has automatically processed and analyzed uploaded data. The result is a set of suggested predictive factors that drive any given variable. Based on the predictive power of the model, estimated by Watson Analytics, we have chosen the most promising ones. Forecasting of stock prices and currencies exchange rates using IBM Watson will be done using IBM Watson analytics "Predict" function in two steps:

1. Choosing factors, which IBM Watson Analytics Suggest as the best predictors

2. Building two factor regression using Ordinary Least Square method in Gretl statistical package

In the Table 12. We can see which variables were chosen as predictors, and which were chosen as targets. Results of applying the IBM Watson analytics predict function are shown in the Appendix 4.

Table 12.

Targets	Input
Prices of stock	Stock Indices & resource prices
12. Exxon Mobil	1. S&P 500
13. Chevron	2. DJI
14. BAC	3. RTS
15. IBM	4. Nikkei
16. P&G	5. CSI
17. Walmart	6. FTSE
18. Apple	7. Shanghai
	8. NASDAQ
	9. Gold
	10. Natural Gas
	11. Brent

Using suggested drivers of predicted values, we have built regression models in Gretl statistical package for each of the observed currency's exchange rate. The results are presented in the Table 13.

Table 13. Description of models built based on IBM Watson.

Model	Model Parameters		Model's Statistic			Predictive Power (%)
	Coefficient	Sig.	R-square	MPE	MAPE	
Exxon Mobil 1			0.712977	-0.10002	2.3478	87.4
const	−4.57266	0.0226				
Gold	0.00725082	<0.0001				
Futsee	0.0339528	<0.0001				
Exxon Mobil 2			0.827575	-0.05769	1.8791	85.2
const	−46.8102	<0.0001				
DJI	0.0405594	<0.0001				
Gold	0.00467673	<0.0001				
Exxon Mobil 3			0.762696	-0.079892	2.2036	83.6
const	−53.5047	<0.0001				
SP 500	0.0398747	<0.0001				
Gold	0.0463417	<0.0001				
Model	Model Parameters		Model's Statistic			Predictive Power (%)
	Coefficient	Sig.	R-square	MPE	MAPE	
IBM 1			0.874521	11.277	11.277	93,00
const	78.6593	<0.0001				
Brent	1.19685	<0.0001				

Model	Coefficient	Sig.	R-square	MPE	MAPE	Predictive Power (%)
NKK225	0.000636722	0.0640				
IBM 2			0.817450	7,5400	7,5400	92,30
const	241.98	<0.0001				
NASDAQ100	0.00448579	0.0916				
USDZAR	−8.25771	<0.0001				
P&G			0.305893	-0.19997	3.199	82.4
const	151.348	<0.0001				
USDJPY	−0.683958	<0.0001				
Brent	0.197202	<0.0001				
Bank of America			0.808380	-0.20033	3.493	90.7
const	−19.479	<0.0001				
Natural Gas	2.11465	<0.0001				
NASDAQ100	0.00683298	<0.0001				
Apple 1			0.928082	-0.066927	2.0298	93,7
const	390.097	<0.0001				
DJI	0.00533233	<0.0001				
USDCNY	−57.7748	<0.0001				
Apple 2			0.934771	1,2012	1,2012	96,30
const	−9.95716	0.0590				
Brent	0.879771	<0.0001				
NASDAQ100	0.0186525	<0.0001				
Walmart 1			0.627249	-1.5062	1,5062	94,50
const	191.434	<0.0001				
USDZAR	−5.50512	<0.0001				
NKK225	−0.00250334	<0.0001				
Walmart 2			0.784290	-1.4992	1,4992	93,30
const	−33.1095	<0.0001				
NKK225	−0.00304983	<0.0001				
Footse100	0.0248156	<0.0001				
Walmart 3			0.601377	-1.7314	1,7314	91,30
const	151.821	<0.0001				
Brent	0.695071	<0.0001				
USDJPY	−0.971299	<0.0001				
Chevron 1			0.754404	-0.29461	4.295	89.3
const	−76.5478	<0.0001				
Gold	0.0393618	<0.0001				
Footse100	0.0192772	<0.0001				

	Model Parameters		Model's Statistic			Predictive Power (%)
Model	Coefficient	Sig.	R-square	MPE	MAPE	
Chevron 2			0.425434	-0.6588	6.4729	88.7
const	119.11	<0.0001				
NASDAQ100	0.0063668	0.0502				
USDZAR	−3.97567	<0.0001				
Coke 1			0.607398	-0.0641	2.0372	79.9
const	43.2875	<0.0001				

Gold	0.00589016	<0.0001				
Natural Gas	−3.42148	<0.0001				
Coke 2			0.407077	7,5483	7,5483	82,3
const	−18.4188	0.0002				
USDCNY	7.84844	<0.0001				
Gold	0.00890342	<0.0001				

Analyzing the results, we can see that three out of thirteen models (IBM 1, IBM 2, and Chevron 2) turned out to be statistically insignificant. That strange result could be explained by the fact that some potentially important predictors were not included in the uploaded dataset. IBM Watson just didn't have enough data to generate good models for these stocks.

R- squared is high or at least tolerable in all cases with the exception for P&G. Additionally, there are two models with a borderline explanatory power – Coke 2 and Chevron 2, R-squared equals 0.407 and 0.425 respectively. Mean percentage errors are quite low, but still they are higher than that of a random walk model.

As a next step, we have estimated potential profitability of trading stocks using regression models, with factors suggested by IBM Watson analytics. For that purpose, we have run the simulation test in Excel 2013. The results are shown in the Table 14. We have used 30 last forecasted values of each currency's exchange rates, for an imitation of real life trading.

Table 14. Results of the Simulation of IBM predictive models.

Trading simulation (Factor regression)		
Model	Factors	Profitability
Apple 1	DJI	0,4529254
	USDCNY	
Apple 2	Brent	0,426352
	Nasdaq 100	
Exxon Mobil 1	Gold	0,0992644
	Futsee 100	
Exxon Mobil 2	Gold	0,4689172
	DJI	
Exxon Mobil 3	SP 500	0,4479888
	Gold	
IBM 1	Brent	0,1694079
	NKK225	
Trading simulation (Factor regression)		
Model	Factors	Profitability
IBM 2	NASDAQ100	0,1816904
	USDZAR	
PG	USD/JPY	-0,0611271
	Brent	
Bank of America	Natural Gas	0,3798725
	NASDAQ100	
Chevron 1	Futsee 100	0,1907649
	Gold	
Chevron 2	NASDAQ100	0,5317324

	USDZAR	
Walmart 1	USDZAR	0,0200946
	NKK 225	
Walmart 2	NKK 225	0,2398046
	Futsee 100	
Walmart 3	Brent	-0,0270778
	USD/JPY	
Coke 1	Gold	0,0003398
	Natural Gas	
Coke 2	Gold	-0,1652693
	USD/CNY	

We have ambivalent results, on one hand; some of the models have demonstrated superior results during the simulation (Apple 1, Apple 2, Exxon Mobil 2, Exxon Mobil 3, and Bank of America), but on the other hand, three models have demonstrated negative result (P&G, Walmart 3, and Coke 2), and one has shown negligibly small profitability (Walmart 1). The lowest results were demonstrated by those models, which turned out to be insignificant (IBM 1, IBM 2). As it was mentioned before, the reason for these results could be absence of some important factors in the dataset.

Overall, IBM Watson generated models have shown results that exceed any other in terms of potential profitability. Average return is 20%, which is way better than that of CAPM. However, there is a problem of separating profitable models from unprofitable ones, and the stability of the desirable performance over the time is still in question.

3.2.3.2 Models for currency's exchange rate forecasting.

We have used free version of IBM Watson analytics to conduct our experiment. After uploading our dataset consisting of 26 variables, IBM Watson Predict option has automatically processed and analyzed uploaded data. The result is a set of suggested predictive factors that drive any given variable. Based on the predictive power of the model, estimated by Watson Analytics, we have chosen the most promising ones. Forecasting of stock prices and currencies exchange rates using IBM Watson will be done using IBM Watson analytics "Predict" function in two steps:

3. Choosing factors, which IBM Watson Analytics Suggest as the best predictors

4. Building two factor regression using Ordinary Least Square method in Gretl statistical package

In the Table 15. We can see which variables were chosen as predictors, and which were chosen as targets. Results of applying the IBM Watson analytics predict function are shown in the Appendix 3.

Table 15. IBM Watson for currencies (source of data: Finam).

Targets	Input	
Currencies	Stock Indices & resource prices	Prices of stock
1. USD/CAD	1. S&P 500	12. Exxon Mobil
2. USD/YEN	2. DJI	13. Chevron
3. USD/ZAR	3. RTS	14. BAC
4. USD/NOR	4. Nikkei	15. IBM
5. USD/CNY	5. CSI	16. P&G
6. USD/RUB	6. FTSE	17. Walmart
7. ERO/USD	7. Shanghai	18. Apple
	8. NASDAQ	
	9. Gold	
	10. Natural Gas	
	11. Brent	

Example: choosing target variables and input variables IBM Watson Analytics displays results, as it shown on the Figure 5.
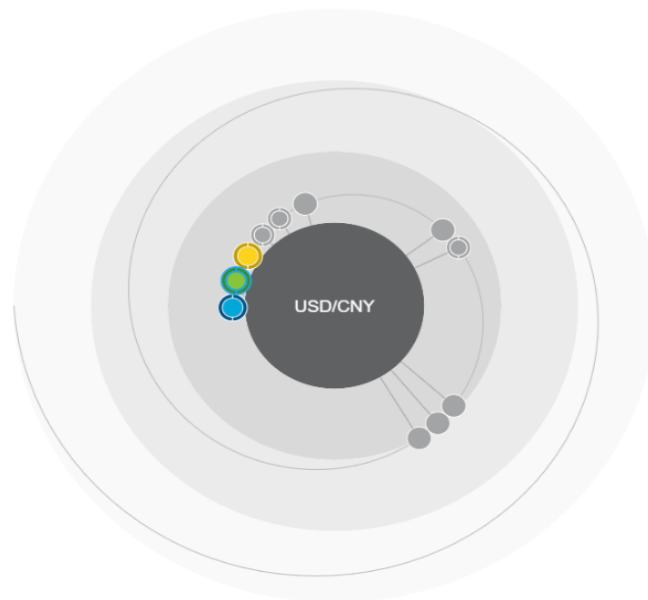
Figure 5. Screenshot of Watson Analytics Predictive function results.

Colored circles represent combination of two predictive factors (Stock Indices, Stock Prices, or Prices of the resources). The closes a circle is to the core, the higher is the predictive power. Using suggested drivers of predicted values, we have built two-factor regression models in Gretl statistical package for each of the observed stock.

Using suggested drivers of predicted values, we have built regression models in Gretl statistical package for each of the observed currency's exchange rate. The results are presented in the Table 16.

Table 16. Description of currency's exchange rate models.

Model	Model Parameters		Model's Statistic			Predictive Power (%)
	Coefficient	Sig,	R-squared	MPE	MAPE	

EUR/USD 1			0,4096	2,8818	2,8818	63,40
const	1,02361	<0,0001				
Gold	0,000299935	<0,0001				
PG	−0,00339848	<0,0001				
USD/CNY 1			0,8981	0,0547	0,0547	96,00
const	6,94919	<0,0001				
Brent	−0,0110162	<0,0001				
Shanghai	−1,60684e-05	0,0335				
USD/CNY 2			0,9198	-0,4768	0,4768	95,70
const	7,22139	<0,0001				
Brent	−0,00990401	<0,0001				
NKK225	−2,03678e-05	<0,0001				
USD/JPY 1			0,8242	-3,1761	3,1761	93,50
const	103,254	<0,0001				
BankAmerica	1,73662	<0,0001				
Gold	−0,00945173	<0,0001				
USD/JPY 2			0,8642	-1,8067	1,8067	92,50
const	109,078	<0,0001				
Gold	−0,020616	<0,0001				
NKK225	0,00184703	<0,0001				
USD/NOK 1			0,8408	-0,9656	0,9656	90,90
const	14,2993	<0,0001				
Gold	−0,00358985	<0,0001				
Natural Gas	−0,776106	<0,0001				
USD/NOK 2			0,8196	-3,5709	3,5709	93,30
const	9,92649	<0,0001				
Natural Gas	−0,0331292	0,0515				
Brent	−0,0323285	<0,0001				
USD/ZAR			0,9076	-3,6134	3,6134	95,60
const	21,3444	<0,0001				
Natural Gas	−0,977925	<0,0001				
Brent	−0,108871	<0,0001				
USD/RUB			0,9182	-5,8449	5,8449	94,50
const	104,251	<0,0001				
Brent	−0,564791	<0,0001				
Shanghai	−0,00349858	<0,0001				

As we can see it in the Table 16, all of the models are statistically significant and have high values of R − squared, with the exception for the Euro to USD exchange rate model. In terms of percentage errors, models still are not capable of beating the Random Walk.

As a next step, we have estimated potential profitability of trading currencies using regression models, with factors suggested by IBM Watson analytics. For that purpose, we have run the simulation test in Excel 2013. We have used 30 last forecasted values of each currency's exchange rates, for an imitation of real life trading.

Results of the simulation tests are shown in the Table 17. In all cases except for Euro to USD, models were able to produce positive results, but the profitability is much lower than that of the stock predicting models (10% vs. 26%), this result is quite surprising. It once again raises question of stability of performance of econometrical models.

Table 17. Results of the simulation for currencies

Trading simulation		
Model	Factors	Profitability
EUR/USD	Gold	0,06876367
	PG	
USD/CNY 1	Brent	-0,0136448
	Shanghai	
USD/CNY 2	Brent	0,00554467
	NKK225	
USD/JPY 1	BankAmerica	0,13238344
	Gold	
USD/JPY 2	Gold	0,12007505
	NKK225	
USD/NOK 1	Gold	0,0160639
	Natural Gas	
USD/NOK 2	Natural Gas	0,08606294
	Brent	
USD/ZAR	Natural Gas	0,12584994
	Brent	
USD/RUB	Brent	0,39969893
	Shanghai	

3.2.3.3 Analysis of the results of stock price forecasting.

We have built a series of predictive models for stock price forecasting and currency exchange rate forecasting. First series were based on the Random Walk model. It was chosen as basis for comparison with other models, as it necessary for any predictive model to outperform random model in order to make at least some sense.

Random walk models have shown unbeatably small deviations of forecasted values from the actual ones, but the random walk model fails to correctly predict the direction of change, therefore it is completely unfit for the purposes of trading. Another type of currency's exchange rate forecasting model we employed is one factor regression, which uses price of the most exported commodity as a predictor. In terms of deviations of forecasts from actual values, they failed to beat the Random Walk, but in terms of potential profitability, as it was demonstrated by

the simulation, they easily outperformed the Random Walk, by demonstrating returns on the level of 20-30%.

Next models we built are CAPM models for the "blue chips" with the index S&P 500 as an average market asset. CAPM has shown poor results in terms of both forecasting accuracy and potential profitability. Its deviation from actual values sometimes exceeded 100%, and only one model has shown substantial returns during the simulation.

Series of stock predictive models based on the suggestions of IBM Watson Analytics have demonstrated results, which are superior to all other models. In terms of forecasting accuracy, they beat all models except for the Random Walk. Additionally, the simulation has demonstrated high returns for most of the suggested models, with the exception for four models with negative and unsubstantial returns. Results of currency's exchange rate forecasting using IBM Watson were worse than that of a simple one-factor regression models, it still beats the Random Walk in potential profitability. It raises the question of spurious correlation between the variables.

Overall, IBM Watson Analytics is capable of suggesting effective predictive models. However, it doesn't provide users with detailed description of the nature of the interdependencies between the variables. It requires further analysis in order to compute actual forecast of the variables in question.

3.3 Conclusion of the Chapter 3.

In the Chapter 3, we have identified four analytical platforms of interest, based on the Gartner's Magic Quadrant for Advanced Analytical Platforms 2016 and Forrester Wave 2015: IBM Watson Analytics, SAS analytics, KNIME, and RapidMiner. The main factor, which has determined such choice, is that they were identified as market leaders, strong performers, and visionaries, with the biggest potential for growth.

We have evaluated the analytical platforms using Analytical Hierarchy process with a set of six KPI's: Visualization, Simplicity of Use, Predictive Analytics Capabilities, Econometric Modeling capabilities, Textual analytics capabilities, and Social Media analytics Capabilities. The results has shown the most preferable analytical platforms for stock price forecasting are IBM Watson and SAS analytics. However, IBM scored a bit better, so we chose it as an analytical platform of choice for stock market forecasting.

Then we have dwelled deeper into how IBM Watson Analytics could be combined with statistical packages. For that purpose, we have built a set of models: theoretically based and those suggested by Watson Analytics. Series of stock predictive models based on the suggestions of IBM Watson Analytics have demonstrated results, which are superior to all other models. In

terms of forecasting accuracy, they beat all models except for the Random Walk. Additionally, the simulation has demonstrated high returns for most of the suggested models, with the exception for four models with negative and unsubstantial returns. Results of currency's exchange rate forecasting using IBM Watson were worse than that of a simple one-factor regression models, it still beats the Random Walk in potential profitability. It raises the question of spurious correlation between the variables.

Overall, IBM Watson Analytics is capable of suggesting effective predictive models. However, it doesn't provide users with detailed description of the nature of the interdependencies between the variables. It requires further analysis in order to compute actual forecast of the variables in question.

Final Conclusions

Discussion of the findings.

In the course of this research, we have completed the set of objectives, stated in the Research Framework Chapter.

First of all, we have evaluated four analytical platforms of interest, based on the Gartner's Magic Quadrant for Advanced Analytical Platforms 2016 and Forrester Wave 2015: IBM Watson Analytics, SAS analytics, KNIME, and RapidMiner. The main factor, which has determined such choice, is that they were identified as market leaders, strong performers, and visionaries, with the biggest potential for growth.

Our next objectives were the evaluation and comparison of the analytical platforms based on their ability to generate predictive models for stock price forecasting.

For the purposes of the evaluation, we have used a set of six KPI's: Visualization, Simplicity of Use, Predictive Analytics Capabilities, Econometric Modeling capabilities, Textual analytics capabilities, and Social Media analytics Capabilities. The result of applying Analytical Hierarchy method has demonstrated that IBM Watson and SAS Analytics are the most appropriate tools, when it comes to forecasting stock market. The whole ranking is shown in the Table 18.

Table 18 . Ranking of Analytical platforms.

Analytical platform	Priority	Rank
IBM Watson	29.4%	1
SAS	29.0 %	2
KNIME	21.6%	3
RapidMiner	20.0%	4

IBM Watson Analytics has won SAS analytics only by a hair. IBM Watson beats SAS at simplicity of use, but SAS wins when it comes to the range of econometrical and statistical tools, which it offers to users. The ability to suggest predictive factors, without preliminary analysis, is what distinguish IBW Watson and SAS from the others. They are superior in their abilities to conduct Predictive Analytics process, while other platforms require statistical expertise in order to use them to full extend.

Our final objectives were to construct, evaluate and compare the results of theoretically based econometric predictive models, and IBM Watson Analytics suggested models. The results has shown that in terms of deviations of forecasts from the actual values of observed variables (measured in terms of Mean Absolute Percentage Errors), the Random Walk is unbeatable. However, when it comes to the potential profitability of the models (assessed trough trading simulation), theoretically based models has shown worse results, that IBM Watson Analytics suggested models, with the exception of the models, based on the prices of most exported commodities. This result could be explained by the fact, that IBM Watson Analytics didn't specify the nature of interdependencies between the variables, meaning that further analysis is required in order to determine the exact econometric equation.

Overall, the effectiveness of IBM Watson Analytics as an effective tool for predictive models suggestion was confirmed.

To sum up, we provide direct answers to the Research questions, as it is shown in the Table 20.

Table 20. Research Questions and answers

Research question	Answer
Which analytical platforms is a better fit for the purposes of stock market forecasting?	IBW Watson Analytics and SAS.
Does IBM Watson Analytics suggest effective predictive models for stock forecasting, in comparison with standard theoretically based econometric models?	Yes, IBM Watson Analytics suggest effective predictive models, however, further analysis is required in order to build the most effective predictive model.

Theoretical implications.

1. Using the theoretical part of this work, similar researches of niche analytical platforms (according to Gartner's Magic quadrant of advanced analytical platforms), such as Prognoz, Accenture, Fico, Megaputer, and Levastorm could be conducted.

2. The research provides a ground for further studies of how different analytical platforms and analytical software tools could be combined in order to construct predictive models.

3. The research can serve as a base for further studies of how big data challenges in financial sector could be tackled using analytical platforms.

4. There are some collateral theoretical results: the theory that currency's exchange rate could be effectively predicted using the price of the mostly exported commodities was confirmed, however, is models have limited applicability, since they could predict exchange rates only for those currencies which are strongly connected to one particular commodity. In other words, it applies only to resource exporting economies.

5. The inability of CAPM to adequately predict stock prices even on the developed stock market was confirmed, therefore the Effective Market Hypothesis is not met on the US stock market.

6. The research has both confirmed and questioned the unbeatable random walk: in terms of the deviation measures, the random walk remains unbeatable, but from the perspective of forecasting the direction of change, it is outperformed by both theoretically based models, and by those that were suggested by IBM Watson Analytics.

Managerial implications.

1. This research provides interested parties (traders) with the recommendations regarding which analytical platforms to use for the purposes of stock price forecasting.

2. The research provides individual traders with tight budget constraints with the no costs combination of analytical platforms (IBM Watson Analytics as a guide, and Statistical Package (Gretl) for the construction of the final model). This combination could prove to be quite effective, since IBM Watson Analytics is the only tool which is capable of suggesting predictive models without preliminary theoretical work.

3. The study has identified the analytical functions, which analytical platform should be able to perform in order to address the business tasks of the financial organizations.

4. The study provides with the criteria, using which analytical platforms can be chosen.

5. The study has contributed to the analysis of the market of financial analytics.

Limitations.

1. Analytical Hierarchy Process imbeds some level of subjectivity: pairwise comparison of the criteria and alternatives could vary depending on the expert.

2. Only four out of many three analytical platforms were chosen.

3. This study was conducted with the use of open source data gathered from the Finam website. Access to the more possible variables harness the possibility for Watson Analytics to generate better predictive models.

4. All predictive models were estimated under the assumption that investor has real time access to all needed information and can react instantly, in accordance with chosen model.

5. Finally, our simulations were run under the assumption that an investor has instant access to all information, needed for the model building, and that an investor can strike deals instantly, before the market reacts on the changes.

List of references

1. (2012, June 26). Small and midsize companies look to make big gains with "big data," according to recent poll conducted on behalf of SAP.

2. Abdullah, L., J. Sunadia, and T. Imran. 2013. Ranking of human capital indicators using analytical hierarchy process. Paper presented on Evaluation of Learning for Performance Improvement International Conference Malaysia, 25 – 26 February, 2013

3. Antweiler, W. and Frank, M.Z. 2004. Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance* 59 (3): 1259-1294.

4. Bologa, A., R., Bologa, and A., Florea. 2010. Big Data and Specific Analysis Methods for Insurance Fraud Detection. *Database Systems Journal* 4 (4): 30-39

5. Cao, M., R., Chychyla, and Stewart, T. 2015. Big Data Analytics in Financial Statement Audit. *Accounting Horizons* 29 (2): 423-429

6. Chung, W. 2014. BizPro: Extracting and categorizing business intelligence factors from textual news articles. *International Journal of Information Management* 34(2): 272–284.

7. Cukier K., 2013. The Economist, Data, data everywhere: A special report on managing information. February 25. Retrieved from http://www.economist.com/node/15557443

8. Curthberston, K. 1996 *Quantitative Financial Economics*. New York. John Willey & Sons Inc.

9. Domenico, F., Kenneth R., and Barbara, R. 2015. Can oil prices forecast exchange rates? An empirical analysis of the relationship between commodity prices and exchange rates. *Journal of International Money and Finance* 54: 116-141

10. Doug, H. Gartner Advanced Analytics Quadrant 2015: Gainers, Losers

11. Earley. E. 2015. Data analytics in auditing: Opportunities and challenges. *Business Horizons* 58: 493—500

12. Elliot, G., and A. Timmermann. 2013. *Handbook of Economic Forecasting*. Elsevier Science and Technology Books, Inc.

13. Fan, J., Han, F., and Liu, H. (2014). Challenges of big data analysis. *National Science Review* 1 (2): 293–314.

14. Finam. http://www.finam.ru/analysis/quotes/?0=&t=8315698

15. Financial Analytics Market - Worldwide Market Forecasts (2013 - 2018). Research and Markets (http://www.researchandmarkets.com/research/6xj66l/financial)

16. Gandomi A., and Murtaza H. 2015. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management* 35: 137–144

17. Gartner IT Glossary (n.d.). Retrieved from http://www.gartner.com/it-glossary/big-data/

18. Gema B., Jason J., and Jung B. 2016. Social big data: Recent achievements and new challenges. *Information Fusion* 28: 45–59

19. IBM corporation website. (https://www.ibm.com/marketplace/cloud/watson-analytics/purchase/us/en-us#product-header-top)

20. Imad, M., Kelly, B. 2014. The unbeatable random walk in exchange rate forecasting: Reality or myth? *Journal of Macroeconomics* 40: 69–81

21. Jagadish H.V. 2015. Big Data and Science: Myths and Reality. *Big Data Research* 2: 49–52

22. Jianzheng L., Jie L., Weifeng L., and Jiansheng W. 2016. Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing* 115: 134–142

23. Joe F., and Hair J. 2007. Knowledge creation in marketing: the role of predictive analytics. *European Business Review*. 19 (4): 303 – 315

24. Johan B., Huina M., and Xiaojun Z. 2015. Twitter mood predicts the stock market *Journal of Computational Science* 2: 1–8.

25. KNIME website. Documentation. (https://tech.knime.org/documentation)

26. Konishi, S., and Kitagawa, G. 2007. Information Criteria and Statistical Modeling *New York: Springer*.

27. Kwan, M. 2014. Big Data's Impact on Trading and Technology. *Journal of Trading* 9 (1): 54-56

28. Kyunghee Y., and Lucas L. 2015. Big Data as Complementary Audit Evidence. *Accounting Horizons* 29: 431-438

29. Mark, E. 2006. Fragile digital data in danger of fading past history's reach. *Atlanta Journal Constitution*, June 7, p. A1.

30. Matlis J. 2006. Predictive Analytics. *Computerworld*, October 9

31. Meese, R., Rogoff, K. 1983. Empirical exchange rate models of the seventies: do they fit out of sample? *Journal of International Economy* 14: 3–24.

32. Microsoft Azure Machine Learning Website. (https://azure.microsoft.com/en-gb/pricing/details/machine-learning/)

33. Min C., and Stewart R. 2015 Big Data Analytics in Financial Statement Audits. *Accounting Horizons*. 29 (2): 423-429

34. Niemira, P.M., and G. F. Zukowski. 1998. *Trading the Fundamentals.* New York. John Willey & Sons Inc.

35. Overview diagram of Microsoft Azure Machine Learning Capabilities. 2016. (https://azure.microsoft.com/en-us/documentation/articles/machine-learning-studio-overview-diagram/)

36. Pozi, S. 2014. Big Data, Big Opportunities. *Best's Reviews.* March.

37. Ramulkan, R. 2015. *Financial Executive.*

38. RapidMiner Studio Manual. 2014. (http://docs.rapidminer.com/downloads/RapidMiner-v6-user-manual.pdf)

39. RapidMiner website. (https://rapidminer.com/products/comparison/)

40. Ruta, D. 2014. Automated trading with machine learning on big data. *IEEE Computer Society*: 824-30

41. SAP. Big Data and Smart Trading: How a Real-time Data Platform Maximizes Trading Opportunities; 2012.

42. SAS Website. Product Documentation. (http://support.sas.com/documentation/index.html)

43. Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., and Tufano, P. 2012. Analytics: The real-world use of big data. How innovative enterprises extract value from uncertain data. IBM Institute for Business Value. Retrieved from http://www-03.ibm.com/systems/hu/resources/the real word use of big data.pdf

44. Schwager, J. D. 1996. *Getting Started in Technical Analysis.* New York. John Willey & Sons Inc.

45. Shmueli, G. 2010. To Explain or to Predict? *Statistical Science* 25 (3): 289-310.

46. Smith, K. 2015. Big Data Discoveries. *Best's Reviews.* November.

47. Spandan G., Soham R., and Satyajit C. 2014. News Analytics and Sentiment Analysis to Predict Stock Price Trends. *International Journal of Computer Science and Information Technologies* 5 (3): 3595-3604

48. Srivastava, U., and Gopalkrishnan S. 2015. Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks. *Procedia Computer Science* 50: 643-52.

49. TechAmerica Foundation's Federal Big Data Commission. 2012. Demystifying big data: A practical guide to transforming the business of Government. Retrieved from http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf

50. Ventana Research. 2016. *Five keys to choosing a comprehensive analytics platform.*

51. Ventana Research. 2016. *Gaining the edge in Banking with Business Analytics.*

52. Ventana Research. 2016. *Perspective Business Analytics.*

53. Wu, H., Harris W., Gongjun Y., Vasudeva Akula C., Jiancheng S. 2015. A novel social media competitive analytics framework with sentiment benchmarks. *Information & Management* 52: 801–812

54. Xinhui T., Rui H., Lei W., Gang L., and Jianfeng Z. 2015. Latency critical big data computing in finance. *The Journal of Finance and Data Science* 1: 33-41

Appendix 1. Specifications of Models.

Model 1: OLS, using observations 2010-02-01:2016-03-21 (T = 321)
Dependent variable: USDCAD

	Coefficient	Std. Error	t-ratio	p-value	
const	1.46591	0.0121235	120.9144	<0.0001	***
Brent	−0.0041492	0.000127235	−32.6105	<0.0001	***

Mean dependent var	1.084380	S.D. dependent var	0.118339
Sum squared resid	1.034060	S.E. of regression	0.056935
R-squared	0.769249	Adjusted R-squared	0.768526
F(1, 319)	1063.443	P-value(F)	1.3e-103
Log-likelihood	465.4615	Akaike criterion	−926.9230
Schwarz criterion	−919.3801	Hannan-Quinn	−923.9113
rho	0.970162	Durbin-Watson	0.050208

Model 2: OLS, using observations 2010-02-01:2016-03-21 (T = 321)
Dependent variable: USDRUB

	Coefficient	Std. Error	t-ratio	p-value	
const	83.0842	1.40507	59.1318	<0.0001	***
Brent	−0.48308	0.014746	−32.7600	<0.0001	***

Mean dependent var	38.66361	S.D. dependent var	13.76334
Sum squared resid	13889.34	S.E. of regression	6.598503
R-squared	0.770869	Adjusted R-squared	0.770151
F(1, 319)	1073.216	P-value(F)	4.4e-104
Log-likelihood	−1060.153	Akaike criterion	2124.305
Schwarz criterion	2131.848	Hannan-Quinn	2127.317
rho	0.959764	Durbin-Watson	0.057043

Model 2: OLS, using observations 2014-11-10:2016-03-21 (T = 72)
Dependent variable: USDNOK

	Coefficient	Std. Error	t-ratio	p-value	
const	10.1061	0.126586	79.8355	<0.0001	***

Brent	−0.039167	0.00237413	−16.4974	<0.0001	***

Mean dependent var	8.067018	S.D. dependent var	0.509242
Sum squared resid	3.766776	S.E. of regression	0.231972
R-squared	0.795420	Adjusted R-squared	0.792498
F(1, 70)	272.1652	P-value(F)	8.06e-26
Log-likelihood	4.052507	Akaike criterion	−4.105013
Schwarz criterion	0.448319	Hannan-Quinn	−2.292318
rho	0.802283	Durbin-Watson	0.378347

Model 1: OLS, using observations 2014-11-10:2016-03-21 (T = 72)
Dependent variable: USDZAR

	Coefficient	Std. Error	t-ratio	p-value	
const	26.4943	3.52805	7.5096	<0.0001	***
GOLD	−0.011423	0.00301772	−3.7853	0.0003	***

Mean dependent var	13.15584	S.D. dependent var	1.608249
Sum squared resid	152.4363	S.E. of regression	1.475690
R-squared	0.169913	Adjusted R-squared	0.158054
F(1, 70)	14.32849	P-value(F)	0.000321
Log-likelihood	−129.1665	Akaike criterion	262.3330
Schwarz criterion	266.8863	Hannan-Quinn	264.1457
rho	0.983043	Durbin-Watson	0.063645

Model 1: OLS, using observations 2015-02-03:2016-03-04 (T = 284)
Dependent variable: BAC

	Coefficient	Std. Error	t-ratio	p-value	
const	0.000181202	0.000738767	0.2453	0.8064	
SP	1.32574	0.0715066	18.5401	<0.0001	***

Mean dependent var	−0.002243	S.D. dependent var	0.018220
Sum squared resid	0.042341	S.E. of regression	0.012253
R-squared	0.549330	Adjusted R-squared	0.547732
F(1, 282)	343.7346	P-value(F)	9.99e-51
Log-likelihood	848.1800	Akaike criterion	−1692.360
Schwarz criterion	−1685.062	Hannan-Quinn	−1689.434
rho	0.060132	Durbin-Watson	1.879509

Model 2: OLS, using observations 2015-02-03:2016-03-04 (T = 284)
Dependent variable: IBM

	Coefficient	Std. Error	t-ratio	p-value
const	−0.00033603	0.000621024	−0.5411	0.5889

3

	Coefficient	Std. Error	t-ratio	p-value	
SP	0.919689	0.06011	15.3001	<0.0001	***

Mean dependent var	−0.002018	S.D. dependent var	0.013910
Sum squared resid	0.029920	S.E. of regression	0.010300
R-squared	0.453587	Adjusted R-squared	0.451649
F(1, 282)	234.0931	P-value(F)	6.87e-39
Log-likelihood	897.4862	Akaike criterion	−1790.972
Schwarz criterion	−1783.674	Hannan-Quinn	−1788.046
rho	0.041344	Durbin-Watson	1.913474

Model 3: OLS, using observations 2015-02-03:2016-03-04 (T = 284)
Dependent variable: MSFT

	Coefficient	Std. Error	t-ratio	p-value	
const	0.00130152	0.000748926	1.7378	0.0833	*
SP	1.13318	0.0724899	15.6322	<0.0001	***

Mean dependent var	−0.000771	S.D. dependent var	0.016941
Sum squared resid	0.043513	S.E. of regression	0.012422
R-squared	0.464251	Adjusted R-squared	0.462351
F(1, 282)	244.3658	P-value(F)	4.22e-40
Log-likelihood	844.3013	Akaike criterion	−1684.603
Schwarz criterion	−1677.305	Hannan-Quinn	−1681.677
rho	0.001750	Durbin-Watson	1.995791

Model 4: OLS, using observations 2015-02-03:2016-03-04 (T = 284)

Dependent variable: PG

	Coefficient	Std. Error	t-ratio	p-value	
const	−0.00079077	0.000448067	−1.7649	0.0787	*
SP	0.671332	0.0433692	15.4795	<0.0001	***

Mean dependent var	−0.002018	S.D. dependent var	0.010090
Sum squared resid	0.015575	S.E. of regression	0.007432
R-squared	0.459370	Adjusted R-squared	0.457453
F(1, 282)	239.6139	P-value(F)	1.52e-39
Log-likelihood	990.1916	Akaike criterion	−1976.383
Schwarz criterion	−1969.085	Hannan-Quinn	−1973.457
rho	0.108871	Durbin-Watson	1.781336

Model 5: OLS, using observations 2015-02-03:2016-03-04 (T = 284)
Dependent variable: Wallmart

	Coefficient	Std. Error	t-ratio	p-value	
const	−0.0015747	0.000696924	−2.2595	0.0246	**

	Coefficient	Std. Error	t-ratio	p-value	
SP	0.592545	0.0674565	8.7841	<0.0001	***

Mean dependent var	−0.002658	S.D. dependent var	0.013022
Sum squared resid	0.037680	S.E. of regression	0.011559
R-squared	0.214836	Adjusted R-squared	0.212051
F(1, 282)	77.16046	P-value(F)	1.56e-16
Log-likelihood	864.7388	Akaike criterion	−1725.478
Schwarz criterion	−1718.180	Hannan-Quinn	−1722.552
rho	0.059662	Durbin-Watson	1.878631

Model 6: OLS, using observations 2015-02-03:2016-03-04 (T = 284)
Dependent variable: Apple

	Coefficient	Std. Error	t-ratio	p-value	
const	−5.80202e-06	0.000709582	−0.0082	0.9935	
SP	1.14417	0.0686817	16.6590	<0.0001	***

Mean dependent var	−0.002098	S.D. dependent var	0.016549
Sum squared resid	0.039062	S.E. of regression	0.011769
R-squared	0.495999	Adjusted R-squared	0.494212
F(1, 282)	277.5229	P-value(F)	7.41e-44
Log-likelihood	859.6272	Akaike criterion	−1715.254
Schwarz criterion	−1707.956	Hannan-Quinn	−1712.329
rho	−0.100408	Durbin-Watson	2.195459

Appendix 2. Specification of models, suggested by Watson Analytics

Model 2: OLS, using observations 2015-01-30:2015-09-10 (T = 224)
Dependent variable: EURUSD

	Coefficient	Std. Error	t-ratio	p-value	
const	2.43518	0.0506615	48.0676	<0.0001	***
Footse100	−8.46547e-05	3.6118e-06	−23.4384	<0.0001	***
USDNOK	−0.0954061	0.00357196	−26.7098	<0.0001	***

Mean dependent var	1.101956	S.D. dependent var	0.023066
Sum squared resid	0.027962	S.E. of regression	0.011248
R-squared	0.764319	Adjusted R-squared	0.762186
F(2, 221)	358.3535	P-value(F)	4.38e-70
Log-likelihood	688.8764	Akaike criterion	−1371.753
Schwarz criterion	−1361.518	Hannan-Quinn	−1367.621
rho	0.811820	Durbin-Watson	0.380038

Model 3: OLS, using observations 2015-01-30:2015-09-10 (T = 224)
Dependent variable: USDJPY

	Coefficient	Std. Error	t-ratio	p-value	
const	111.292	2.97802	37.3710	<0.0001	***
BankAmerica	1.72864	0.062178	27.8014	<0.0001	***
Coke	−0.451838	0.0582075	−7.7625	<0.0001	***

Mean dependent var	120.0950	S.D. dependent var	3.329135
Sum squared resid	379.9253	S.E. of regression	1.311152
R-squared	0.846280	Adjusted R-squared	0.844889
F(2, 221)	608.3391	P-value(F)	1.36e-90
Log-likelihood	−377.0150	Akaike criterion	760.0301
Schwarz criterion	770.2650	Hannan-Quinn	764.1614
rho	0.883048	Durbin-Watson	0.230867

Model 4: OLS, using observations 2015-01-30:2015-09-10 (T = 224)
Dependent variable: USDJPY

	Coefficient	Std. Error	t-ratio	p-value	
const	103.254	2.96398	34.8363	<0.0001	***
BankAmerica	1.73662	0.0716346	24.2427	<0.0001	***
Gold	−0.00945173	0.00189369	−4.9912	<0.0001	***

Mean dependent var	120.0950	S.D. dependent var	3.329135
Sum squared resid	434.5323	S.E. of regression	1.402216
R-squared	0.824186	Adjusted R-squared	0.822595
F(2, 221)	518.0038	P-value(F)	3.79e-84
Log-likelihood	−392.0561	Akaike criterion	790.1123
Schwarz criterion	800.3472	Hannan-Quinn	794.2436
rho	0.917690	Durbin-Watson	0.176774

Model 5: OLS, using observations 2015-01-30:2015-09-10 (T = 224)
Dependent variable: USDJPY

	Coefficient	Std. Error	t-ratio	p-value	
const	109.078	2.38948	45.6492	<0.0001	***
Gold	−0.020616	0.00148216	−13.9094	<0.0001	***
NKK225	0.00184703	6.42593e-05	28.7434	<0.0001	***

Mean dependent var	120.0950	S.D. dependent var	3.329135
Sum squared resid	335.5763	S.E. of regression	1.232252
R-squared	0.864224	Adjusted R-squared	0.862995
F(2, 221)	703.3392	P-value(F)	1.50e-96
Log-likelihood	−363.1130	Akaike criterion	732.2260
Schwarz criterion	742.4610	Hannan-Quinn	736.3573
rho	0.862122	Durbin-Watson	0.271155

Dependent variable: USDNOK

	Coefficient	Std. Error	t-ratio	p-value

	Coefficient	Std. Error	t-ratio	p-value	
const	11.834	0.175394	67.4711	<0.0001	***
Gold	−0.00182426	0.000159866	−11.4112	<0.0001	***
Brent	−0.0299062	0.000884068	−33.8280	<0.0001	***

Mean dependent var	8.245108	S.D. dependent var	0.389453
Sum squared resid	3.847638	S.E. of regression	0.131947
R-squared	0.886243	Adjusted R-squared	0.885213
F(2, 221)	860.8666	P-value(F)	4.8e-105
Log-likelihood	137.3467	Akaike criterion	−268.6934
Schwarz criterion	−258.4584	Hannan-Quinn	−264.5620
rho	0.845472	Durbin-Watson	0.303042

Model 7: OLS, using observations 2015-01-30:2015-09-10 (T = 224)
Dependent variable: USDNOK

	Coefficient	Std. Error	t-ratio	p-value	
const	22.7357	0.356379	63.7964	<0.0001	***
EURUSD	−8.0025	0.299609	−26.7098	<0.0001	***
Footse100	−0.000878508	1.79426e-05	−48.9621	<0.0001	***

Mean dependent var	8.245108	S.D. dependent var	0.389453
Sum squared resid	2.345393	S.E. of regression	0.103018
R-squared	0.930657	Adjusted R-squared	0.930030
F(2, 221)	1483.035	P-value(F)	8.5e-129
Log-likelihood	192.7874	Akaike criterion	−379.5748
Schwarz criterion	−369.3398	Hannan-Quinn	−375.4434
rho	0.789141	Durbin-Watson	0.421785

Model 8: OLS, using observations 2015-01-30:2015-09-10 (T = 224)
Dependent variable: USDRUB

	Coefficient	Std. Error	t-ratio	p-value	
const	104.251	0.960429	108.5464	<0.0001	***
Brent	−0.564791	0.0220999	−25.5563	<0.0001	***
Shanghai	−0.00349858	0.000409167	−8.5505	<0.0001	***

Mean dependent var	63.87961	S.D. dependent var	7.947528
Sum squared resid	1152.483	S.E. of regression	2.283606
R-squared	0.918179	Adjusted R-squared	0.917438
F(2, 221)	1240.007	P-value(F)	7.4e-121
Log-likelihood	−501.3014	Akaike criterion	1008.603
Schwarz criterion	1018.838	Hannan-Quinn	1012.734
rho	0.903221	Durbin-Watson	0.178160

Model 9: OLS, using observations 2015-01-30:2015-09-10 (T = 224)
Dependent variable: USDRUB

	Coefficient	Std. Error	t-ratio	p-value

	Coefficient	Std. Error	t-ratio	p-value	
const	118.242	1.30709	90.4620	<0.0001	***
Brent	−0.398387	0.0219356	−18.1617	<0.0001	***
RTSI	−0.0406336	0.00247756	−16.4007	<0.0001	***

Mean dependent var	63.87961	S.D. dependent var	7.947528
Sum squared resid	691.7760	S.E. of regression	1.769239
R-squared	0.950887	Adjusted R-squared	0.950443
F(2, 221)	2139.413	P-value(F)	2.4e-145
Log-likelihood	−444.1352	Akaike criterion	894.2705
Schwarz criterion	904.5054	Hannan-Quinn	898.4018
rho	0.860782	Durbin-Watson	0.275189

Model 10: OLS, using observations 2015-01-30:2015-09-10 (T = 224)
Dependent variable: USDZAR

	Coefficient	Std. Error	t-ratio	p-value	
const	21.3444	0.213486	99.9806	<0.0001	***
Brent	−0.108871	0.00500692	−21.7442	<0.0001	***
NaturalGas	−0.977925	0.143414	−6.8189	<0.0001	***

Mean dependent var	13.55606	S.D. dependent var	1.533969
Sum squared resid	48.50850	S.E. of regression	0.468503
R-squared	0.907556	Adjusted R-squared	0.906719
F(2, 221)	1084.815	P-value(F)	5.4e-115
Log-likelihood	−146.4926	Akaike criterion	298.9853
Schwarz criterion	309.2202	Hannan-Quinn	303.1166
rho	0.879994	Durbin-Watson	0.205604

Model 11: OLS, using observations 2015-01-30:2015-09-10 (T = 224)
Dependent variable: USDZAR

	Coefficient	Std. Error	t-ratio	p-value	
const	19.7559	0.68409	28.8792	<0.0001	***
Brent	−0.137358	0.00344813	−39.8356	<0.0001	***
Gold	0.000517846	0.000623526	0.8305	0.4071	

Mean dependent var	13.55606	S.D. dependent var	1.533969
Sum squared resid	58.53171	S.E. of regression	0.514635
R-squared	0.888454	Adjusted R-squared	0.887445
F(2, 221)	880.1244	P-value(F)	5.5e-106
Log-likelihood	−167.5296	Akaike criterion	341.0591
Schwarz criterion	351.2940	Hannan-Quinn	345.1904
rho	0.898736	Durbin-Watson	0.164108

Model 15: OLS, using observations 2015-01-30:2015-09-10 (T = 224)
Dependent variable: IBM

	Coefficient	Std. Error	t-ratio	p-value	
const	78.6593	5.23793	15.0173	<0.0001	***
Brent	1.19685	0.043626	27.4344	<0.0001	***
NKK225	0.000636722	0.000342023	1.8616	0.0640	*

Mean dependent var	149.9098	S.D. dependent var	14.20635
Sum squared resid	5647.317	S.E. of regression	5.055044
R-squared	0.874521	Adjusted R-squared	0.873385
F(2, 221)	770.1224	P-value(F)	2.5e-100
Log-likelihood	−679.2987	Akaike criterion	1364.597
Schwarz criterion	1374.832	Hannan-Quinn	1368.729
rho	0.917691	Durbin-Watson	0.194187

Model 16: OLS, using observations 2015-01-30:2015-09-10 (T = 224)
Dependent variable: IBM

	Coefficient	Std. Error	t-ratio	p-value	
const	241.98	13.0694	18.5150	<0.0001	***
NASDAQ100	0.00448579	0.00264733	1.6945	0.0916	*
USDZAR	−8.25771	0.273081	−30.2391	<0.0001	***

Mean dependent var	149.9098	S.D. dependent var	14.20635
Sum squared resid	8215.836	S.E. of regression	6.097190
R-squared	0.817450	Adjusted R-squared	0.815798
F(2, 221)	494.8132	P-value(F)	2.41e-82
Log-likelihood	−721.2856	Akaike criterion	1448.571
Schwarz criterion	1458.806	Hannan-Quinn	1452.702
rho	0.938997	Durbin-Watson	0.120422

Model 22: OLS, using observations 2015-01-30:2015-09-10 (T = 224)
Dependent variable: Apple

	Coefficient	Std. Error	t-ratio	p-value	
const	−9.95716	5.24672	−1.8978	0.0590	*
Brent	0.879771	0.0182226	48.2790	<0.0001	***
NASDAQ100	0.0186525	0.00122047	15.2831	<0.0001	***

Mean dependent var	116.2154	S.D. dependent var	10.84963
Sum squared resid	1712.286	S.E. of regression	2.783505
R-squared	0.934771	Adjusted R-squared	0.934181
F(2, 221)	1583.529	P-value(F)	9.9e-132
Log-likelihood	−545.6434	Akaike criterion	1097.287
Schwarz criterion	1107.522	Hannan-Quinn	1101.418
rho	0.829285	Durbin-Watson	0.335655

	Coefficient	Std. Error	t-ratio	p-value	
const	191.434	9.74956	19.6351	<0.0001	***

	Coefficient	Std. Error	t-ratio	p-value	
USDZAR	−5.50512	0.303515	−18.1379	<0.0001	***
NKK225	−0.00250334	0.000344423	−7.2682	<0.0001	***

Mean dependent var	69.57540	S.D. dependent var	8.425007
Sum squared resid	5900.174	S.E. of regression	5.166974
R-squared	0.627249	Adjusted R-squared	0.623875
F(2, 221)	185.9442	P-value(F)	4.38e-48
Log-likelihood	−684.2044	Akaike criterion	1374.409
Schwarz criterion	1384.644	Hannan-Quinn	1378.540
rho	0.953880	Durbin-Watson	0.091654

Model 24: OLS, using observations 2015-01-30:2015-09-10 (T = 224)
Dependent variable: Wallmart

	Coefficient	Std. Error	t-ratio	p-value	
const	−33.1095	4.52621	−7.3151	<0.0001	***
NKK225	−0.00304983	0.000261806	−11.6492	<0.0001	***
Footse100	0.0248156	0.000918855	27.0071	<0.0001	***

Mean dependent var	69.57540	S.D. dependent var	8.425007
Sum squared resid	3414.405	S.E. of regression	3.930623
R-squared	0.784290	Adjusted R-squared	0.782338
F(2, 221)	401.7626	P-value(F)	2.47e-74
Log-likelihood	−622.9428	Akaike criterion	1251.886
Schwarz criterion	1262.121	Hannan-Quinn	1256.017
rho	0.903227	Durbin-Watson	0.187717

Model 25: OLS, using observations 2015-01-30:2015-09-10 (T = 224)
Dependent variable: Wallmart

	Coefficient	Std. Error	t-ratio	p-value	
const	151.821	13.7927	11.0074	<0.0001	***
Brent	0.695071	0.0380797	18.2531	<0.0001	***
USDJPY	−0.971299	0.121221	−8.0126	<0.0001	***

Mean dependent var	69.57540	S.D. dependent var	8.425007
Sum squared resid	6309.686	S.E. of regression	5.343278
R-squared	0.601377	Adjusted R-squared	0.597770
F(2, 221)	166.7043	P-value(F)	7.28e-45
Log-likelihood	−691.7201	Akaike criterion	1389.440
Schwarz criterion	1399.675	Hannan-Quinn	1393.571
rho	0.952624	Durbin-Watson	0.078185

Model 3: OLS, using observations 1960-01-01:1960-08-11 (T = 224)
Dependent variable: Coke

	Coefficient	Std. Error	t-ratio	p-value

	Coefficient	Std. Error	t-ratio	p-value	
const	−18.4188	4.88511	−3.7704	0.0002	***
USDCNY	7.84844	0.668813	11.7349	<0.0001	***
Gold	0.00890342	0.00149152	5.9693	<0.0001	***

Mean dependent var	41.68634	S.D. dependent var	1.654961
Sum squared resid	362.1419	S.E. of regression	1.280098
R-squared	0.407077	Adjusted R-squared	0.401711
F(2, 221)	75.86486	P-value(F)	8.25e-26
Log-likelihood	−371.6459	Akaike criterion	749.2918
Schwarz criterion	759.5268	Hannan-Quinn	753.4232
rho	0.950880	Durbin-Watson	0.121149

Model 4: OLS, using observations 1960-01-01:1960-08-11 (T = 224)
Dependent variable: EURUSD

	Coefficient	Std. Error	t-ratio	p-value	
const	1.02361	0.0261078	39.2069	<0.0001	***
Gold	0.000299935	2.50379e-05	11.9793	<0.0001	***
PG	−0.00339848	0.00035657	−9.5310	<0.0001	***

Mean dependent var	1.101956	S.D. dependent var	0.023066
Sum squared resid	0.070045	S.E. of regression	0.017803
R-squared	0.409610	Adjusted R-squared	0.404267
F(2, 221)	76.66453	P-value(F)	5.14e-26
Log-likelihood	586.0265	Akaike criterion	−1166.053
Schwarz criterion	−1155.818	Hannan-Quinn	−1161.922
rho	0.883499	Durbin-Watson	0.238517

Model 6: OLS, using observations 1960-01-01:1960-08-11 (T = 224)
Dependent variable: USDCNY

	Coefficient	Std. Error	t-ratio	p-value	
const	6.94919	0.0176275	394.2246	<0.0001	***
Brent	−0.0110162	0.000405616	−27.1592	<0.0001	***
Shanghai	−1.60684e-05	7.50977e-06	−2.1397	0.0335	**

Mean dependent var	6.346921	S.D. dependent var	0.130698
Sum squared resid	0.388227	S.E. of regression	0.041913
R-squared	0.898085	Adjusted R-squared	0.897162
F(2, 221)	973.7338	P-value(F)	2.6e-110
Log-likelihood	394.2327	Akaike criterion	−782.4655
Schwarz criterion	−772.2305	Hannan-Quinn	−778.3341
rho	0.880092	Durbin-Watson	0.229048

Model 7: OLS, using observations 1960-01-01:1960-08-11 (T = 224)
Dependent variable: USDCNY

	Coefficient	Std. Error	t-ratio	p-value	
const	7.22139	0.0385345	187.4007	<0.0001	***
Brent	−0.00990401	0.000320948	−30.8586	<0.0001	***
NKK225	−2.03678e-05	2.5162e-06	−8.0946	<0.0001	***

Mean dependent var	6.346921	S.D. dependent var	0.130698
Sum squared resid	0.305649	S.E. of regression	0.037189
R-squared	0.919763	Adjusted R-squared	0.919037
F(2, 221)	1266.665	P-value(F)	8.6e-122
Log-likelihood	421.0179	Akaike criterion	−836.0358
Schwarz criterion	−825.8008	Hannan-Quinn	−831.9045
rho	0.844110	Durbin-Watson	0.288563

Model 8: OLS, using observations 1960-01-01:1960-08-11 (T = 224)
Dependent variable: USDJPY

	Coefficient	Std. Error	t-ratio	p-value	
const	103.254	2.96398	34.8363	<0.0001	***
BankAmerica	1.73662	0.0716346	24.2427	<0.0001	***
Gold	−0.00945173	0.00189369	−4.9912	<0.0001	***

Mean dependent var	120.0950	S.D. dependent var	3.329135
Sum squared resid	434.5323	S.E. of regression	1.402216
R-squared	0.824186	Adjusted R-squared	0.822595
F(2, 221)	518.0038	P-value(F)	3.79e-84
Log-likelihood	−392.0561	Akaike criterion	790.1123
Schwarz criterion	800.3472	Hannan-Quinn	794.2436
rho	0.917690	Durbin-Watson	0.176774

Model 9: OLS, using observations 1960-01-01:1960-08-11 (T = 224)
Dependent variable: USDJPY

	Coefficient	Std. Error	t-ratio	p-value	
const	109.078	2.38948	45.6492	<0.0001	***
Gold	−0.020616	0.00148216	−13.9094	<0.0001	***
NKK225	0.00184703	6.42593e-05	28.7434	<0.0001	***

Mean dependent var	120.0950	S.D. dependent var	3.329135
Sum squared resid	335.5763	S.E. of regression	1.232252
R-squared	0.864224	Adjusted R-squared	0.862995
F(2, 221)	703.3392	P-value(F)	1.50e-96
Log-likelihood	−363.1130	Akaike criterion	732.2260
Schwarz criterion	742.4610	Hannan-Quinn	736.3573
rho	0.862122	Durbin-Watson	0.271155

Model 10: OLS, using observations 1960-01-01:1960-08-11 (T = 224)
Dependent variable: USDNOK

	Coefficient	Std. Error	t-ratio	p-value	
const	14.2993	0.2173	65.8042	<0.0001	***
Gold	−0.00358985	0.00017834	−20.1293	<0.0001	***
NaturalGas	−0.776106	0.0282488	−27.4739	<0.0001	***

Mean dependent var	8.245108	S.D. dependent var	0.389453
Sum squared resid	5.383486	S.E. of regression	0.156076
R-squared	0.840835	Adjusted R-squared	0.839394
F(2, 221)	583.7466	P-value(F)	6.37e-89
Log-likelihood	99.72848	Akaike criterion	−193.4570
Schwarz criterion	−183.2220	Hannan-Quinn	−189.3256
rho	0.835213	Durbin-Watson	0.328522

Model 11: OLS, using observations 1960-01-01:1960-08-11 (T = 224)
Dependent variable: USDNOK

	Coefficient	Std. Error	t-ratio	p-value	
const	9.92649	0.0757235	131.0886	<0.0001	***
NaturalGas	−0.0331292	0.0508692	−0.6513	0.5156	
Brent	−0.0323285	0.00177596	−18.2034	<0.0001	***

Mean dependent var	8.245108	S.D. dependent var	0.389453
Sum squared resid	6.102984	S.E. of regression	0.166179
R-squared	0.819562	Adjusted R-squared	0.817929
F(2, 221)	501.8999	P-value(F)	6.67e-83
Log-likelihood	85.67901	Akaike criterion	−165.3580
Schwarz criterion	−155.1231	Hannan-Quinn	−161.2267
rho	0.888607	Durbin-Watson	0.208145

Model 12: OLS, using observations 1960-01-01:1960-08-11 (T = 224)
Dependent variable: USDZAR

	Coefficient	Std. Error	t-ratio	p-value	
const	21.3444	0.213486	99.9806	<0.0001	***
NaturalGas	−0.977925	0.143414	−6.8189	<0.0001	***
Brent	−0.108871	0.00500692	−21.7442	<0.0001	***

Mean dependent var	13.55606	S.D. dependent var	1.533969
Sum squared resid	48.50850	S.E. of regression	0.468503
R-squared	0.907556	Adjusted R-squared	0.906719
F(2, 221)	1084.815	P-value(F)	5.4e-115
Log-likelihood	−146.4926	Akaike criterion	298.9853
Schwarz criterion	309.2202	Hannan-Quinn	303.1166
rho	0.879994	Durbin-Watson	0.205604

Model 14: OLS, using observations 1960-01-01:1960-08-11 (T = 224)
Dependent variable: USDRUB

	Coefficient	Std. Error	t-ratio	p-value	
const	104.251	0.960429	108.5464	<0.0001	***
Brent	−0.564791	0.0220999	−25.5563	<0.0001	***
Shanghai	−0.00349858	0.000409167	−8.5505	<0.0001	***

Mean dependent var	63.87961	S.D. dependent var	7.947528
Sum squared resid	1152.483	S.E. of regression	2.283606
R-squared	0.918179	Adjusted R-squared	0.917438
F(2, 221)	1240.007	P-value(F)	7.4e-121
Log-likelihood	−501.3014	Akaike criterion	1008.603
Schwarz criterion	1018.838	Hannan-Quinn	1012.734
rho	0.903221	Durbin-Watson	0.178160

Model 2: OLS, using observations 1960-01-01:1960-09-19 (T = 225)
Dependent variable: Apple

	Coefficient	Std. Error	t-ratio	p-value	
const	390.097	20.0778	19.4293	<0.0001	***
DJI	0.00533233	0.000430637	12.3824	<0.0001	***
USDCNY	−57.7748	2.19356	−26.3384	<0.0001	***

Mean dependent var	116.1877	S.D. dependent var	10.83334
Sum squared resid	1890.653	S.E. of regression	2.918297
R-squared	0.928082	Adjusted R-squared	0.927434
F(2, 222)	1432.419	P-value(F)	1.3e-127
Log-likelihood	−558.7261	Akaike criterion	1123.452
Schwarz criterion	1133.700	Hannan-Quinn	1127.588
rho	0.830495	Durbin-Watson	0.333046

Model 1: OLS, using observations 1960-01-01:1961-03-27 (T = 322)
Dependent variable: USDCAD

	Coefficient	Std. Error	t-ratio	p-value	
const	1.4658	0.0120031	122.1180	<0.0001	***
Brent	−0.00414812	0.000126136	−32.8862	<0.0001	***

Mean dependent var	1.085059	S.D. dependent var	0.118781
Sum squared resid	1.034076	S.E. of regression	0.056846
R-squared	0.771673	Adjusted R-squared	0.770960
F(1, 320)	1081.499	P-value(F)	1.2e-104
Log-likelihood	467.4097	Akaike criterion	−930.8195
Schwarz criterion	−923.2704	Hannan-Quinn	−927.8056
rho	0.969058	Durbin-Watson	0.051482

Model 3: OLS, using observations 1960-01-01:1960-11-10 (T = 225)
Dependent variable: ExxonMobil

	Coefficient	Std. Error	t-ratio	p-value	
const	−4.57266	3.76875	−1.2133	0.2263	
Footse100	0.00725082	0.000467742	15.5017	<0.0001	***
Gold	0.0339528	0.00306931	11.0621	<0.0001	***

Mean dependent var	81.49080	S.D. dependent var	4.686620
Sum squared resid	1412.159	S.E. of regression	2.522117
R-squared	0.712977	Adjusted R-squared	0.710392
F(2, 222)	275.7292	P-value(F)	6.74e-61
Log-likelihood	−525.8983	Akaike criterion	1057.797
Schwarz criterion	1068.045	Hannan-Quinn	1061.933
rho	0.884229	Durbin-Watson	0.232393

Model 4: OLS, using observations 1960-01-01:1960-11-10 (T = 225)
Dependent variable: ExxonMobil

	Coefficient	Std. Error	t-ratio	p-value	
const	−46.8102	3.93945	−11.8824	<0.0001	***
Gold	0.0405594	0.00226864	17.8782	<0.0001	***
DJI	0.00467673	0.00019986	23.4000	<0.0001	***

Mean dependent var	81.49080	S.D. dependent var	4.686620
Sum squared resid	848.3353	S.E. of regression	1.954822
R-squared	0.827575	Adjusted R-squared	0.826022
F(2, 222)	532.7585	P-value(F)	1.83e-85
Log-likelihood	−468.5684	Akaike criterion	943.1368
Schwarz criterion	953.3851	Hannan-Quinn	947.2731
rho	0.861211	Durbin-Watson	0.281813

Model 5: OLS, using observations 1960-01-01:1960-11-10 (T = 225)
Dependent variable: ExxonMobil

	Coefficient	Std. Error	t-ratio	p-value	
const	−53.5047	5.14368	−10.4020	<0.0001	***
SP500	0.0398747	0.00217159	18.3620	<0.0001	***
Gold	0.0463417	0.00262328	17.6655	<0.0001	***

Mean dependent var	81.49080	S.D. dependent var	4.686620
Sum squared resid	1167.540	S.E. of regression	2.293292
R-squared	0.762696	Adjusted R-squared	0.760559
F(2, 222)	356.7554	P-value(F)	4.56e-70
Log-likelihood	−504.4985	Akaike criterion	1014.997
Schwarz criterion	1025.245	Hannan-Quinn	1019.133
rho	0.893820	Durbin-Watson	0.215242

Model 9: OLS, using observations 1960-01-01:1960-11-10 (T = 225)
Dependent variable: PG

	Coefficient	Std. Error	t-ratio	p-value	
const	151.348	8.77858	17.2406	<0.0001	***
USDJPY	−0.683958	0.0772603	−8.8526	<0.0001	***
Brent	0.197202	0.0245568	8.0304	<0.0001	***

Mean dependent var	78.98476	S.D. dependent var	4.117686
Sum squared resid	2636.217	S.E. of regression	3.445990

R-squared	0.305893	Adjusted R-squared	0.299639	
F(2, 222)	48.91763	P-value(F)	2.50e-18	
Log-likelihood	−596.1236	Akaike criterion	1198.247	
Schwarz criterion	1208.496	Hannan-Quinn	1202.384	
rho	0.949963	Durbin-Watson	0.097331	

Model 10: OLS, using observations 1960-01-01:1960-11-10 (T = 225)
Dependent variable: BankAmerica

	Coefficient	Std. Error	t-ratio	p-value	
const	−19.479	1.29258	−15.0699	<0.0001	***
NaturalGas	2.11465	0.123832	17.0768	<0.0001	***
NASDAQ100	0.00683298	0.000290473	23.5236	<0.0001	***

Mean dependent var	15.97807	S.D. dependent var	1.554276
Sum squared resid	103.6922	S.E. of regression	0.683434
R-squared	0.808380	Adjusted R-squared	0.806653
F(2, 222)	468.2703	P-value(F)	2.24e-80
Log-likelihood	−232.1104	Akaike criterion	470.2209
Schwarz criterion	480.4692	Hannan-Quinn	474.3571
rho	0.908326	Durbin-Watson	0.212804

Model 11: OLS, using observations 1960-01-01:1960-11-10 (T = 225)
Dependent variable: Chevron

	Coefficient	Std. Error	t-ratio	p-value	
const	−76.5478	7.28786	−10.5035	<0.0001	***
Gold	0.0393618	0.0059353	6.6318	<0.0001	***
Footse100	0.0192772	0.000904501	21.3125	<0.0001	***

Mean dependent var	93.40311	S.D. dependent var	9.797375
Sum squared resid	5280.668	S.E. of regression	4.877170
R-squared	0.754404	Adjusted R-squared	0.752191
F(2, 222)	340.9617	P-value(F)	2.06e-68
Log-likelihood	−674.2783	Akaike criterion	1354.557
Schwarz criterion	1364.805	Hannan-Quinn	1358.693
rho	0.938045	Durbin-Watson	0.126310

Model 12: OLS, using observations 1960-01-01:1960-11-10 (T = 225)
Dependent variable: Chevron

	Coefficient	Std. Error	t-ratio	p-value	
const	119.11	15.9567	7.4646	<0.0001	***
NASDAQ100	0.0063668	0.00323398	1.9687	0.0502	*
USDZAR	−3.97567	0.333502	−11.9210	<0.0001	***

|			
Mean dependent var	93.40311	S.D. dependent var	9.797375
Sum squared resid	12353.99	S.E. of regression	7.459799
R-squared	0.425434	Adjusted R-squared	0.420258
F(2, 222)	82.18939	P-value(F)	1.94e-27
Log-likelihood	−769.8950	Akaike criterion	1545.790

Schwarz criterion	1556.038	Hannan-Quinn	1549.926
rho	0.979301	Durbin-Watson	0.042312

Model 16: OLS, using observations 1960-01-01:1960-11-10 (T = 225)
Dependent variable: Coke

	Coefficient	Std. Error	t-ratio	p-value	
const	43.2875	1.47685	29.3108	<0.0001	***
Gold	0.00589016	0.00120946	4.8701	<0.0001	***
NaturalGas	−3.42148	0.191359	−17.8799	<0.0001	***

Mean dependent var	41.70929	S.D. dependent var	1.686764
Sum squared resid	250.2125	S.E. of regression	1.061642
R-squared	0.607398	Adjusted R-squared	0.603861
F(2, 222)	171.7292	P-value(F)	8.49e-46
Log-likelihood	−331.2098	Akaike criterion	668.4196
Schwarz criterion	678.6679	Hannan-Quinn	672.5559
rho	0.908895	Durbin-Watson	0.211163

Appendix 3. Results of the IBM Watson Analytics Predict function for currencies.
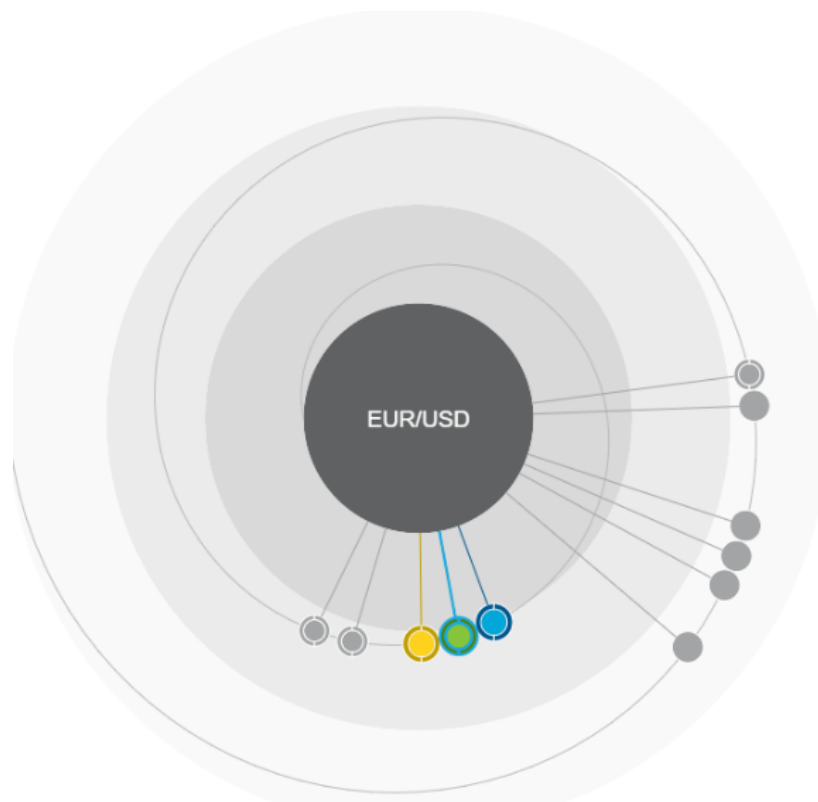
Blue: P&G and NKK225
Green: P&G and Gold.
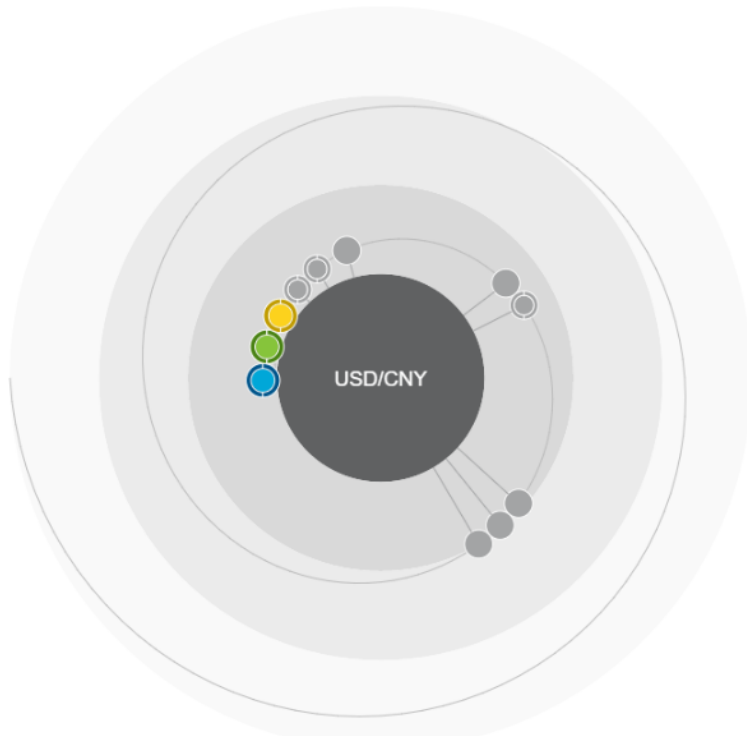
Blue: Brent and Shanghai
Green: Brent and NKK 225
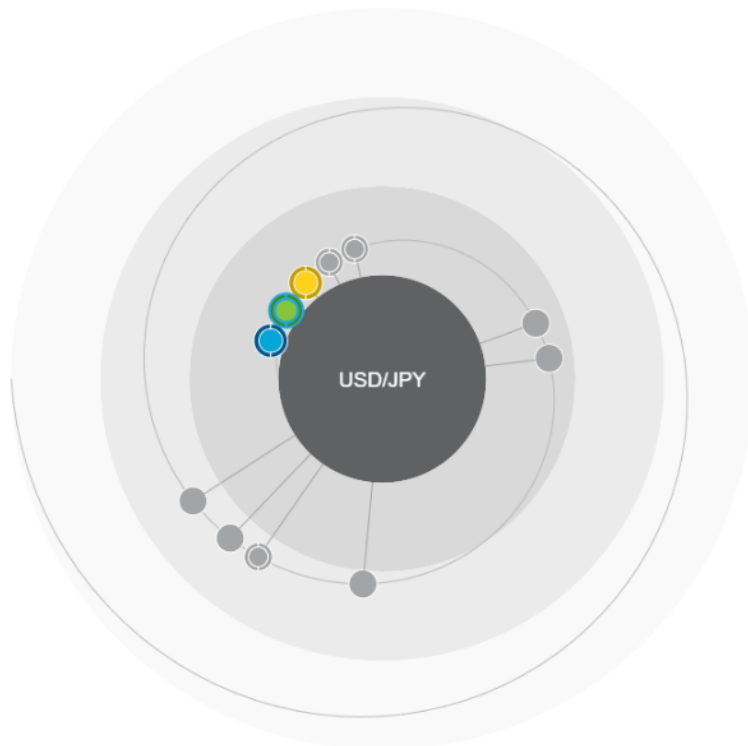
Blue: Bank America and Coke
Green: Bank America and Gold
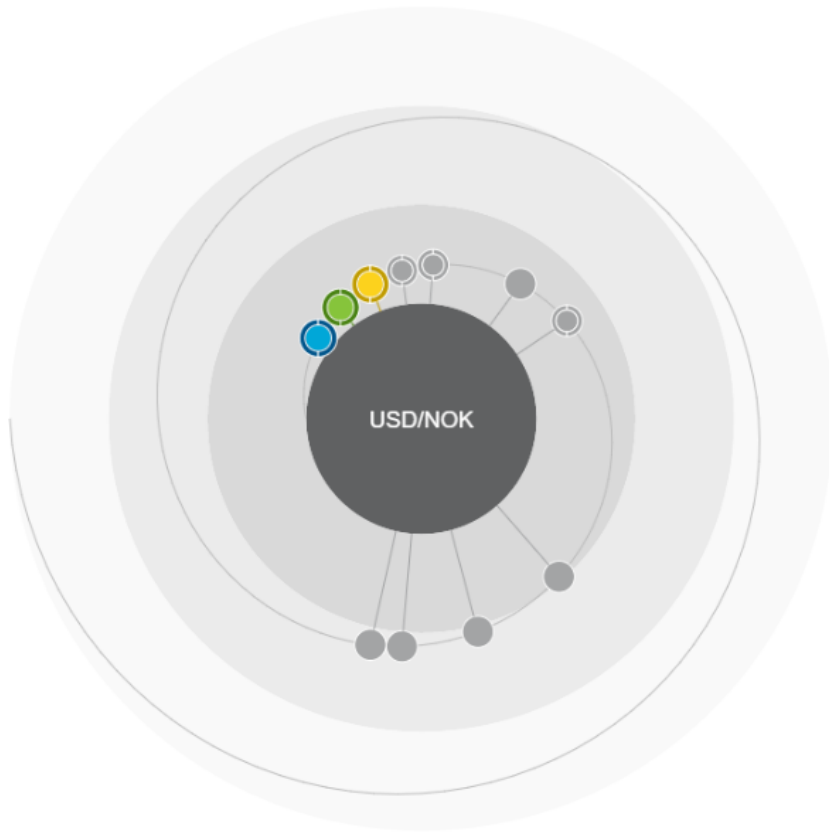
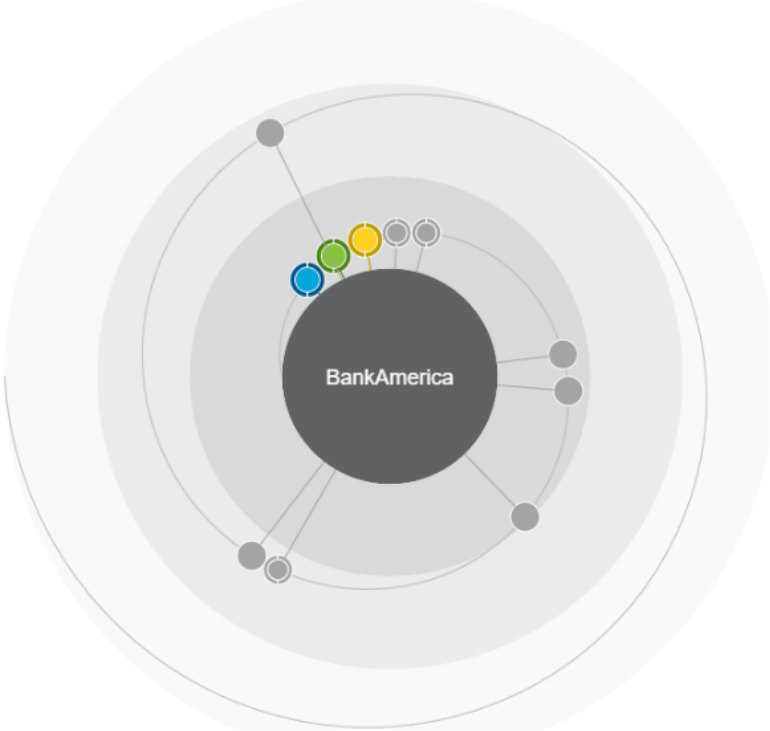Blue: Brent and Gold
Green: Bank Gold and Natural Gas
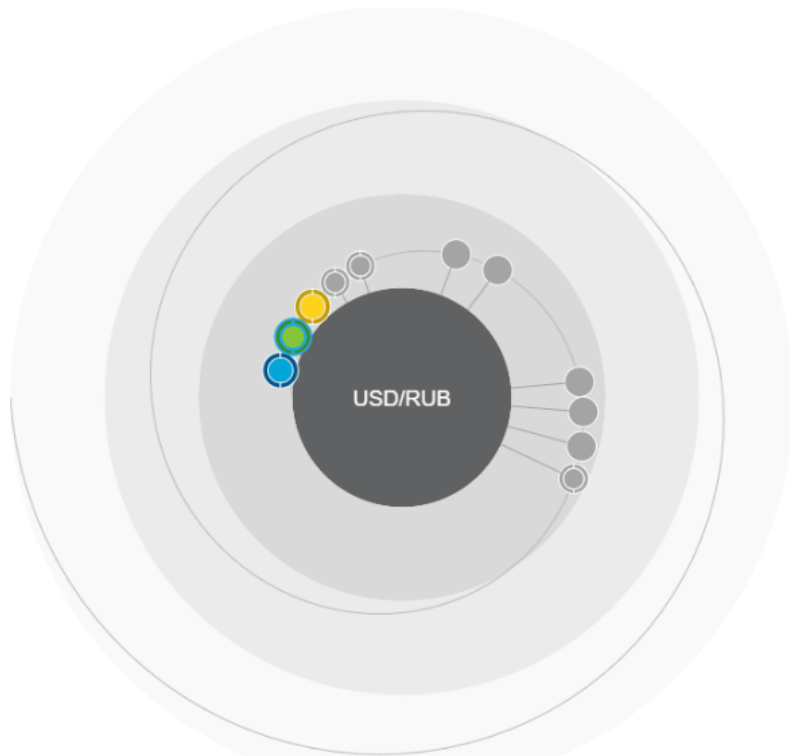
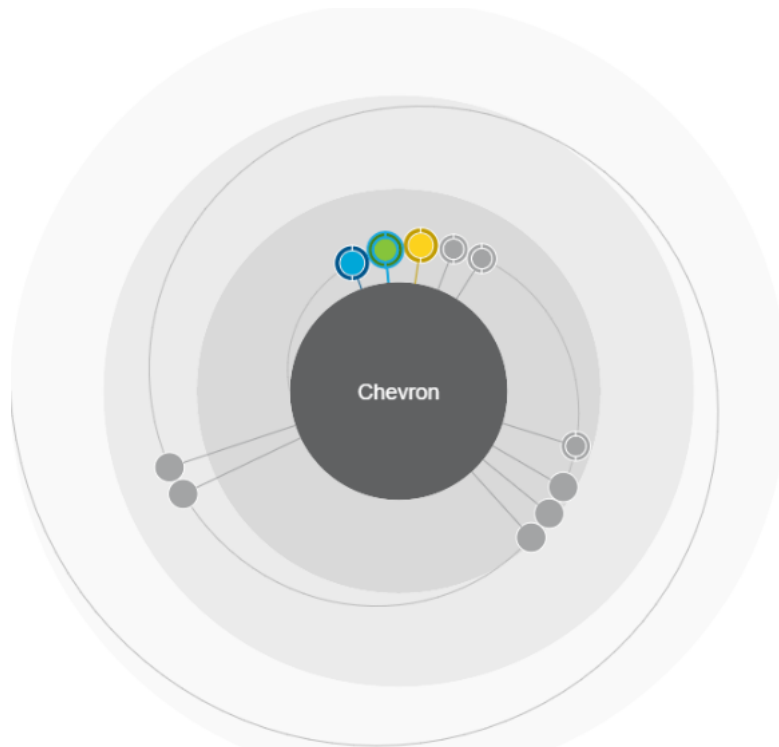Blue: Brent and Shanghai
Green: Brent and CSI300

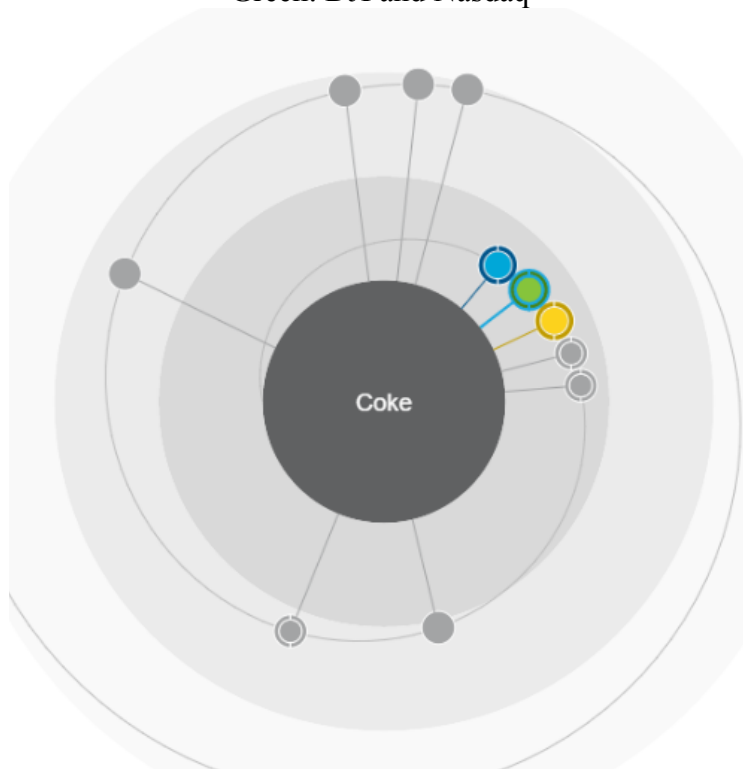**Appendix 4. Results of the IBM Watson Analytics Predict function for stocks.**
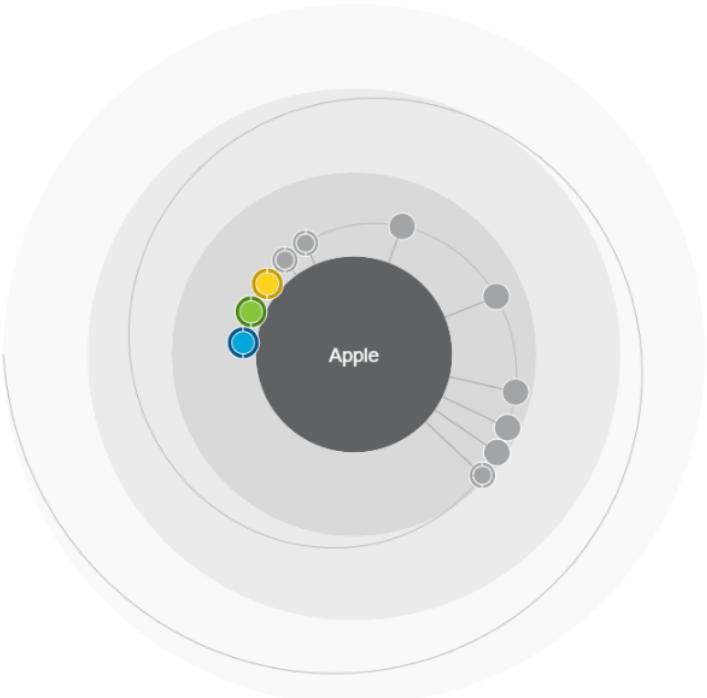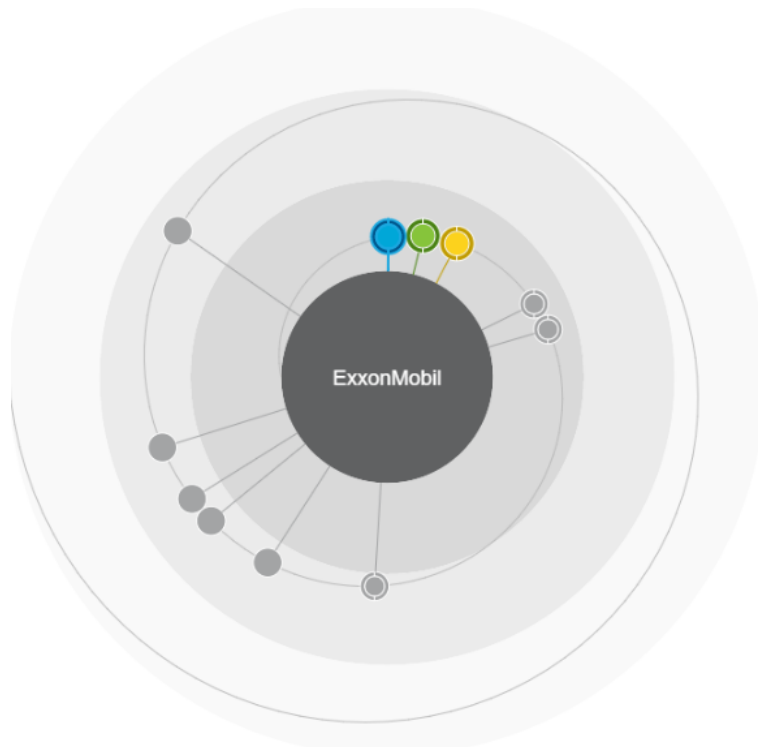

Blue: Brent and Nasdaq
Green NKK225 and Brent

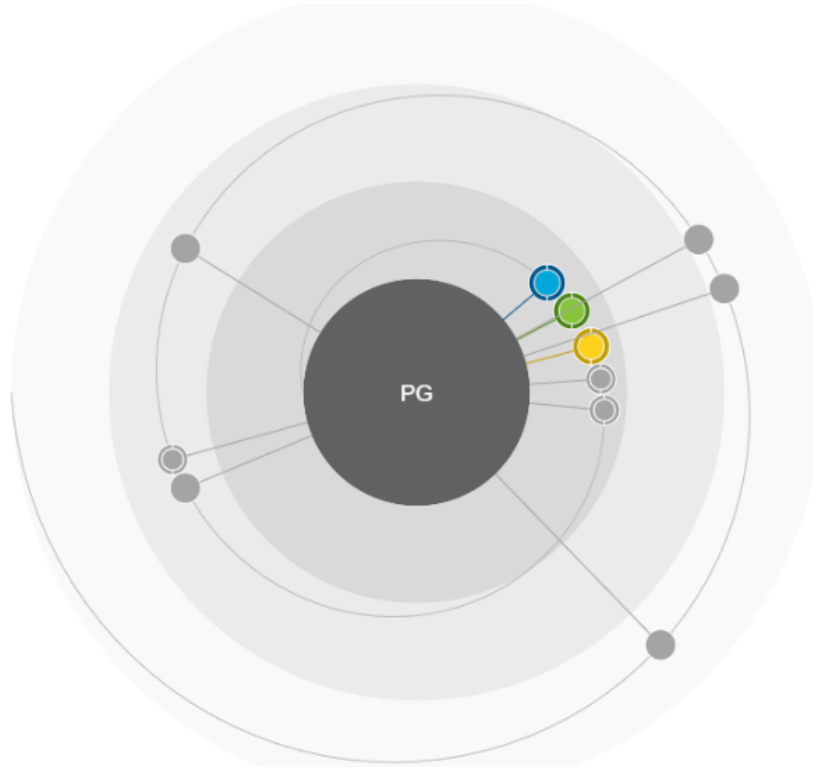Blue: Nasdaq and Natural Gas
Green Shanghai and Natural Gas

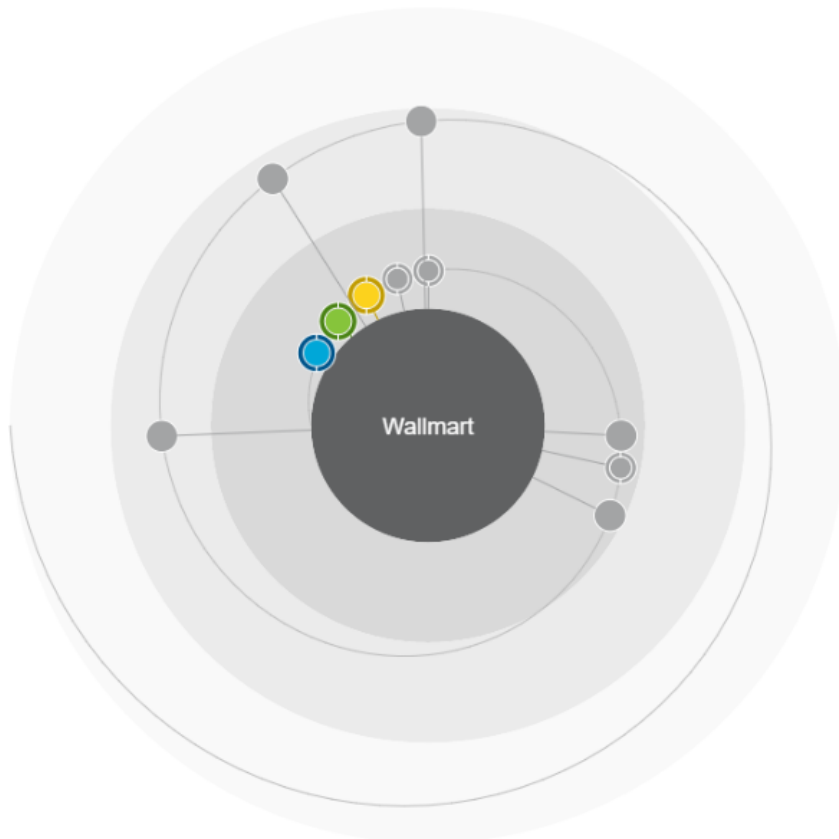Blue: Futsee 100 and Gold
Green: DJI and Nasdaq

Blue: Gold and Natural Gas
Green: Futsee and Natural Gas

Blue: Futsee and Gold
Green: DJI and Gold

Blue: Gold and Brent
Green: Brent and NKK 225

Blue: Futsee and NKK 225
Green: Futsee and Nasdaq